

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068645 A2(51) International Patent Classification⁷: C12N 15/12,
I/21, C07K 14/47, 16/18, C12Q 1/68, G01N 33/574,
A61K 38/17, 31/7088, 39/395Lane, Apartment 218, San Jose, CA 95134 (US). SUN,
Yongming [US/US]; 551 Shoal Drive, Redwood City, CA
94065 (CN). LIU, Chenghua [CN/US]; 1125 Ranchero
Way, #14, San Jose, CA 95117 (US). TURNER, Leah, R.
[US/US]; 939 Rosette Court, Sunnyvale, CA 94086 (US).

(21) International Application Number: PCT/US01/45151

(22) International Filing Date:
20 November 2001 (20.11.2001)(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(26) Publication Language: English

(30) Priority Data:
60/249,992 20 November 2000 (20.11.2000) US(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:
US 60/249,992 (CIP)
Filed on 20 November 2000 (20.11.2000)(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).(71) Applicant (*for all designated States except US*): DI-
ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).

(72) Inventors; and

Published:

(75) Inventors/Applicants (*for US only*): SALCEDA, Susana
[AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136
(US). MACINA, Roberto, A. [AR/US]; 4118 Crescendo
Avenue, San Jose, CA 95136 (US). RECIPON, Herve
[FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115
(US). CAFFERKEY, Robert [IE/US]; 350 Elan Village— without international search report and to be republished
upon receipt of that reportFor two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/068645 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

AC

DEX-050J

**COMPOSITIONS AND METHODS
RELATING TO BREAST SPECIFIC GENES AND PROTEINS**

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/249,992 filed November 20, 2000, which is herein incorporated by
reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
polypeptides present in normal and neoplastic breast cells, including fragments, variants
10 and derivatives of the nucleic acids and polypeptides. The present invention also relates
to antibodies to the polypeptides of the invention, as well as agonists and antagonists of
the polypeptides of the invention. The invention also relates to compositions comprising
the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists
of the invention and methods for the use of these compositions. These uses include
15 identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-
cancerous disease states in breast tissue, identifying breast tissue and monitoring and
identifying and/or designing agonists and antagonists of polypeptides of the invention.
The uses also include gene therapy, production of transgenic animals and cells, and
production of engineered breast tissue for treatment and research.

20 **BACKGROUND OF THE INVENTION**

Excluding skin cancer, breast cancer, also called mammary tumor, is the most
common cancer among women, accounting for a third of the cancers diagnosed in the
United States. One in nine women will develop breast cancer in her lifetime and about
192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths.
25 Bevers, *Primary Prevention of Breast Cancer*, in *BREAST CANCER*, 20-54 (Kelly K Hunt
et al., ed., 2001); Kochanek et al., 49 Nat'l. Vital Statistics Reports 1, 14 (2001).

In the treatment of breast cancer, there is considerable emphasis on detection and
risk assessment because early and accurate staging of breast cancer has a significant
impact on survival. For example, breast cancer detected at an early stage (stage T0,
30 discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not
detected until a late stage (i.e., stage T4), the five-year survival rate is reduced to 13%.
AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998).

-2-

Some detection techniques, such as mammography and biopsy, involve increased discomfort, expense, and/or radiation, and are only prescribed only to patients with an increased risk of breast cancer.

Current methods for predicting or detecting breast cancer risk are not optimal.

- 5 One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regimens for high risk patients. A patient's risk of breast cancer has been positively associated with increasing age, nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast
- 10 disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer
- 15 limited their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than the risk that comes with growing older. NIH Publication No. 00-1556 (2000).

- Current screening methods for detecting cancer, such as breast self exam, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent
- 20 their widespread adoption. Breast self exams, while useful, are unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by palpitation. Ultrasound measurements require skilled operators at an increased expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation
- 25 used in mammography because prior chest radiation is a factor associated with an increase incidence of breast cancer.

- At this time, there are no adequate methods of breast cancer prevention. The current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy
- 30 before cancer diagnosis) which are drastic measures that limit their adoption even among women with increased risk of breast cancer. Bevers, *supra*.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., 249 JAMA

-3-

- 1881 (1983)) MUC-1 (Frische and Liu, 22 J. Clin. Ligand 320 (2000)), HER-2/neu (Haris et al., 15 Proc.Am.Soc.Clin.Oncology. A96 (1996)), uPA, PAI-1, LPA, LPC, RAK and BRCA (Esteve and Fritsche, *Serum and Tissue Markers for Breast Cancer*, in BREAST CANCER, 286-308 (2001)). These markers have problems with limited
- 5 sensitivity, low correlation, and false negatives which limit their use for initial diagnosis. For example, while the BRCA1 gene mutation is useful as an indicator of an increased risk for breast cancer, it has limited use in cancer diagnosis because only 6.2 % of breast cancers are BRCA1 positive. Malone et al., 279 JAMA 922 (1998). *See also*, Mewman et al., 279 JAMA 915 (1998) (correlation of only 3.3%).
- 10 Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ,
- 15 intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 is characterized by
- 20 tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T4. Within stage T4, T4a indicates extension of the tumor to the chest wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. AJCC Cancer Staging Handbook pp. 159-70
- 25 (Irvin D. Fleming et al. eds., 5th ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., 7 Breast Cancer Research and Treatment 147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., 90 J.Nat'l.Cancer Inst. 1346 (1998); Paik et al., 90 J.Nat'l.Cancer Inst. 1361 (1998);
- 30 Hutchins et al., 17 Proc.Am.Soc.Clin.Oncology A2 (1998).; and Simpson et al., 18 J.Clin.Oncology 2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be

-4-

assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or to other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., 18 J. Clin. Oncology 2059 (2000). Generally, pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could differentiate between different stages of invasion. Progress in this field will allow more rapid and reliable method for treating breast cancer patients.

Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., 127 Annals of Internal Medicine 1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for early stage breast cancer. Fisher et al., 16 J. of Clinical Oncology 441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may be employed. These treatment methods are equally effective in the early stages of breast cancer.

Patients with stage I and stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and stage IV patients. Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent restaging because cancer

-5-

is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, ¶. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998).

In an effort to provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence which could be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page et al., 75 Cancer 1219 (1995); Fisher et al., 75 Cancer 1223 (1995); Silverstein et al., 77 Cancer 2267 (1996).

As discussed above, each of the methods for diagnosing and staging breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of breast cancers to optimize treatment methods. Finally, there is a need for sensitive molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of breast cancers following remission.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast; identify and monitor breast tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered breast tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to breast cells and/or breast tissue. These breast specific nucleic acids

-6-

(BSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the BSNA is genomic DNA, then the BSNA is a breast specific gene (BSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 116 through 218. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 115. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a BSP, or that selectively hybridize or exhibit substantial sequence similarity to a BSNA, as well as allelic variants of a nucleic acid molecule encoding a BSP, and allelic variants of a BSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a BSP or that comprises a part of a nucleic acid sequence of a BSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a BSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a BSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a BSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a BSNA.

Another object of the invention is to provide methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a BSP.

The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a BSP.

-7-

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

- 5 Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and
- 10 treating breast cancer and non-cancerous disease states in breast. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring breast tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered breast tissue for treatment and research.
- 15 The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast. The invention provides methods of using the polypeptides of the invention to identify and/or monitor breast tissue, and to produce engineered breast tissue.
- 20 The agonists and antagonists of the instant invention may be used to treat breast cancer and non-cancerous disease states in breast and to produce engineered breast tissue.
- Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for
- 25 comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

- Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by
- 30 those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture,

-8-

molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various

5 general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing

10 Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of

15 which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and

20 pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

25 A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and

30 "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both

naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

-10-

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other

nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded
5 deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for
10 use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are
15 synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are
20 not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such
25 as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-
30 occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate,

phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, e.g., the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000);

Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other

higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example,

-16-

when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

5
$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/\text{N}),$$

wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using
10 mismatched probes, pools of degenerate probes or “guessmers,” as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term “digestion” or “digestion of DNA” refers to catalytic cleavage of the
15 DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the
20 purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are
25 ordinarily used, but conditions may vary in accordance with standard procedures, the supplier’s instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

30 The term “ligation” refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs.

Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to
5 transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is
10 identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one
15 exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic
20 acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing
25 Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that
30 nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at

any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a BSP or is a BSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

5 The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g., Leung et al., Technique 1: 11-15 (1989) and Caldwell et al., PCR Methods Applic. 2: 28-33 (1992).*

10 The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g., Reidhaar-Olson et al., Science 241: 53-57 (1988).*

 The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR
15 reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

 The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random
20 fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A. 91: 10747-10751 (1994).* DNA shuffling can be carried out between several related genes ("Family shuffling").

 The term "*in vivo* mutagenesis" refers to a process of generating random
25 mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

30 The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

-20-

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial
5 artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors
10 can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are
15 often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is
20 intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term
25 "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode
30 the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

-21-

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

5 The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance
10 with the invention comprises a BSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

 The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins
15 from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using
20 protein purification techniques well-known in the art.

 A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%,
25 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using
30 HPLC or other means well-known in the art for purification.

 The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a

-22-

contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40
5 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such
10 modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of
15 cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with
20 radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and
25 antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

30 The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein

-23-

comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs.

The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus

sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

- 5 A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or
- 10 deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least
- 15 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity
- 20 may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

- Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example,
- 25 single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions
- 30 are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent

-25-

sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which
5 are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such
10 as -, -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other
15 similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism
20 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"
25 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence
30 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

-26-

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions.

- 5 A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative
10 substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

- For instance, the following six groups each contain amino acids that are
15 conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 20 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

- Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is
25 any change having a nonnegative value in the PAM250 log-likelihood matrix.

- Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid
30 substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from

-27-

different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer
 5 program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

	Expectation value:	10 (default)
	Filter:	seg (default)
10	Cost to open a gap:	11 (default)
	Cost to extend a gap:	1 (default)
	Max. alignments:	100 (default)
	Word size:	11 (default)
	No. of descriptions:	100 (default)
15	Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number
 20 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best
 25 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

30 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of

-28-

intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the

5 polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al.,*

10 *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to

15 that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston *et al., Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific

20 antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al., Structure* 2: 1121-1123 (1994). One or more

25 CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody

30 that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or

5 "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that
10 purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that
15 normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side
20 chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than $1\ \mu\text{M}$, preferably less than $100\ \text{nM}$ and most preferably less than $10\ \text{nM}$.

The term "patient" as used herein includes human and veterinary subjects.

25 Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "breast specific" refers to a nucleic acid molecule or polypeptide that is
30 expressed predominantly in the breast as compared to other tissues in the body. In a preferred embodiment, a "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "breast specific" nucleic acid molecule or polypeptide is

expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the breast or to breast cells or tissue or that are derived from such nucleic acid molecules. These isolated breast specific nucleic acids (BSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast, a breast-specific polypeptide (BSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 116 through 218. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 115.

A BSNA may be derived from a human or from another animal. In a preferred embodiment, the BSNA is derived from a human or other mammal. In a more preferred embodiment, the BSNA is derived from a human or other primate. In an even more preferred embodiment, the BSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a BSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a BSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a BSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 116 through 218. In an even more preferred embodiment,

-31-

the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 115.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 116 through 218. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 115. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a BSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human BSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 218. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a BSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 218; more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a BSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a BSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a BSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115. In a preferred
5 embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a BSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 115, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a BSNA, more preferably at least 95%,
10 more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a BSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one
15 that exhibits sequence identity over its entire length to a BSNA or to a nucleic acid molecule encoding a BSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the BSNA or the nucleic acid molecule encoding a BSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at
20 least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 116 through 218 or demonstrates
25 significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 115. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the BSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated
30 species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring

substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is
5 experimentally produced by directed mutation of a BSNA. Further, the substantially similar nucleic acid molecule may or may not be a BSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a BSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a BSNA or a nucleic acid encoding a BSP. For instance, single nucleotide polymorphisms
10 (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide
15 polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes
20 a BSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a BSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1
25 through 115. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide,
30 and may or may not encode a polypeptide that is a BSP. However, in a preferred embodiment, the part encodes a BSP. In one aspect, the invention comprises a part of a BSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a BSNA. In a third aspect, the

invention comprises a part of a nucleic acid molecule that is an allelic variant of a BSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a BSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides.

- 5 The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutin or a polypeptide analog, as described below.

- 10 Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides
15 of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

- In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide
20 analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression
25 of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

- In a preferred embodiment, isolated nucleic acid molecules can include nucleotide
30 analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

-35-

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as ^{-32}P -dATP, ^{-32}P -dCTP, ^{-32}P -dGTP, ^{-32}P -dTTP, ^{-32}P -3'dATP, ^{-32}P -ATP, ^{-32}P -CTP, ^{-32}P -GTP, ^{-32}P -UTP, ^{-35}S -dATP, α - ^{35}S -GTP, α - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See Henegariu et al., Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3'

-36-

hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids.

For example, fluorophores can be attached using a cisplatin reagent that reacts with the
5 N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA
to provide a stable coordination complex between the nucleic acid and fluorophore label
(Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA
and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes,
Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994);
10 Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference.
As another example, nucleic acids can be labeled using a disulfide-containing linker
(FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or
thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a
free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or
15 other marker.

One or more independent or interacting labels can be incorporated into the
nucleic acid molecules of the present invention. For example, both a fluorophore and a
moiety that in proximity thereto acts to quench fluorescence can be included to report
specific hybridization through release of fluorescence quenching or to report
20 exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996);
Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci.
USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras
et al., *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517;
5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280
25 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids
Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by
reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more
native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside
30 bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in
Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied
Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.),
Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd

(1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See Gamper et al., Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

- 5 Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, 10 thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 15 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside 20 linkages may be used for antisense techniques.

- Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include 25 those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; 30 and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;

-38-

5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr.*

Opin. Struct. Biol. 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

25 *Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a BSNA, such as

deletions, insertions, translocations, and duplications of the BSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g., Andreeff et al.* (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated
5 herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be
10 restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify BSNA in, and isolate BSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated
15 nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g., Schvartzacher et al., In Situ Hybridization*, Springer-Verlag New York (2000), the disclosure of which
20 is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level
25 characterization of mRNAs that hybridize to BSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and
30 are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a BSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 116 through 218. In another preferred embodiment, the probe or primer is derived from a BSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115.

10 In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer
15 in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g., Sambrook et al., 1989, supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide
20 probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic
25 Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of
30 which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g., in* Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert

(ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

5 PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

10 The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

15 Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

20 Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

25 In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another
30 embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene,

polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

- 5 The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound,
- 10 hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is,
- 15 therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

- Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in
- 20 which such vectors have been introduced.

- The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes
- 25 (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

- 30 Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems:

Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are
5 incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that
10 expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or
15 GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

20 In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial
25 hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host,
30 selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

-45-

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIp1ac, YEplac and YCplac).

Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSFTTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will

-46-

typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A).

- 5 Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in
- 10 mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus),

15 and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

- 20 It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of
- 25 the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control

30 sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include

-47-

splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within
5 particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7
10 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon,
15 Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast α -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the
20 transcription termination signals from the *CYC1* or *ADH1* gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the
25 Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the BSNA of interest. Often, expression is enhanced by
30 incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may
5 alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are
10 described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors
15 include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the
20 *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible
25 because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

30 In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for

-49-

example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. See Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and

-50-

Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.* Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.* Heim

-51-

et al., *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)).

Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*,

- 5 Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference
- 10 in their entireties. *See also* Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

- Fusions to the IgG Fc region increase serum half life of protein pharmaceutical
- 15 products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

- For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable
- 20 expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin
- 25 C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The *bsd* gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

- Replication incompetent retroviral vectors, typically derived from Moloney
- 30 murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories,

-52-

Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide BSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquitination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-

-53-

anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

- 5 General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/ResearchCommittees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and
- 10 <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).
- 15

- Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational
- 20 modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is
- 25 phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue.
- 30 Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance,
5 the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of
10 amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide
15 cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to
20 which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

25 Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies
30 specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined

by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification.

10 For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

15 desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a

20 nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may

25 delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be

30 considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the

-56-

product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

5 The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such
10 polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

 Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic
15 acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

 Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral
20 infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral
25 expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

30 A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture.

-57-

Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from

5 *Spodoptera frugiperda*, e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3

10 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic

15 Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from breast are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human breast cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further

20 details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

25 Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be

30 packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by

-58-

treatment, e.g., with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold®
5 Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth
10 in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the
15 action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed
20 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective
25 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension
30 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by

-59-

using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be
5 coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™
10 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).
Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) ([http://www.bio-rad.com/LifeScience/pdf/](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)
15 [New_Gene_Pulser.pdf](http://www.bio-rad.com/LifeScience/pdf/New_Gene_Pulser.pdf)); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA*
20 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid
25 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A
Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.),
30 Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

-60-

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate
5 fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a
10 breast specific polypeptide (BSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 116 through 218. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification
15 and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a BSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 116
20 through 218. A polypeptide that comprises only a fragment of an entire BSP may or may not be a polypeptide that is also a BSP. For instance, a full-length polypeptide may be breast-specific, while a fragment thereof may be found in other tissues as well as in breast. A polypeptide that is not a BSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing
25 animals to prepare anti-BSP antibodies. However, in a preferred embodiment, the part or fragment is a BSP. Methods of determining whether a polypeptide is a BSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. See, e.g., Geysen *et al.*, *Proc. Natl. Acad. Sci. USA*
30 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by

native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.,* Lerner, *Nature* 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a BSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a

-62-

fragment of polypeptide of the invention, preferably a BSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a BSP, in a host cell.

5 By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion,
10 duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be breast-specific. In a preferred embodiment, the mutein is breast-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion,
15 duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 116 through 218. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218. In yet a
20 more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218.

A mutein may be produced by isolation from a naturally-occurring mutant cell,
25 tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell
30 comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is breast-specific, as described below. Multiple random mutations can be introduced into the

5 gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*,

10 Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the

15 polypeptide is homologous to a BSP. In an even more preferred embodiment, the polypeptide is homologous to a BSP selected from the group having an amino acid sequence of SEQ ID NO: 116 through 218. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a BSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity

20 to an comprising an amino acid sequence of SEQ ID NO: 116 through 218. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218. In a yet more

25 preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%,

30 99.7%, 99.8% or 99.9% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a BSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a BSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSNA is selected from the group consisting of SEQ ID NO: 1 through 115. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a BSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSP is selected from the group consisting of SEQ ID NO: 116 through 218.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 116 through 218. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the BSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a BSP. Further, the homologous protein may or may not encode polypeptide that is a BSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a BSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an

antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a BSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 218. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 115.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 218, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are

available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

5 It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched
10 circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and
15 such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-
20 reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430,
25 Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa
30 Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY

-67-

- 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine,
- 5 Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS,
- 10 DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI,
- 15 SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

- 20 The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

- 25 The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-BSP antibodies.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 30 usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

-68-

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 5 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 116 through 218. In a preferred embodiment, the analog is one that comprises one or
- 10 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a BSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- 15 --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a BSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 20 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 25 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000);
- 30 Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added
5 using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and *t*BOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during
10 synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be
15 incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side
20 chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-
25 aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic
30 acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranilic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-

-70-

aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 116 through 218, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid

-71-

molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 115, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115.

- 5 The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino
10 acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

- The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger
15 polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

- As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention
20 can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain
25 their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

- As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of
30 recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a

-72-

glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein
5 encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*,
15 *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbriozio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that
20 functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A*. 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95,
25 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl*
30 *Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

-73-

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

- 5 The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

- Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST,
10 immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast μ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of
15 IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a
20 Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the BSP.

- As further described below, the isolated polypeptides, muteins, fusion proteins,
25 homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize BSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly BSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser
30 scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of BSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of BSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

-75-

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a BSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a

-76-

polypeptide that comprises SEQ ID NO: 116 through 218, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a BSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or *visa versa*. In addition, alternative splice forms of a BSP may be indicative of cancer. Differential degradation of the C or N-terminus of a BSP may also be a marker or target for anticancer therapy. For example, a BSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-BSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human breast.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune

-77-

immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that
5 express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299;
10 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the
15 antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be
20 obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are
25 hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids
30 of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate
5 chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual,
10 Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by
15 reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in
20 immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge *et al.*, *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

25 Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential
30 Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entirety, and thus need not be detailed here.

-79-

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable

regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

5 For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. See, e.g., Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 10 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. See, e.g., Li *et al.*, *Protein Expr. Purif.* 15 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can 20 also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein 25 by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. See, e.g. Pollock *et al.*, *J. Immunol Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 30 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated
5 by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic
10 animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present
15 invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin.*
20 *Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively
25 inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more
30 suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.,* United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco *et al.*, *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan *et al.*, *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding

-83-

of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

- 5 For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose
10 oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN);
15 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

- Other substrates can be used to produce products for local deposition that are
20 luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic
25 compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g., Thorpe et al., Methods Enzymol.* 133: 331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such
30 enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

-84-

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the
5 antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such
10 as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY
15 R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red
20 (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting
25 applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb ,
30 ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI

-85-

contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

5 The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin
10 Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind
15 specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present
20 invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

25 For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

30 As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma

-86-

cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention,
5 to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

10 In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular
15 application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid
20 molecule encoding a BSP. In a preferred embodiment, the BSP comprises an amino acid sequence selected from SEQ ID NO: 116 through 218, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a BSNA of the invention, preferably a BSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID
25 NO: 1 through 115, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human BSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-
30 human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g., Hogan et al., Manipulating the Mouse Embryo: A Laboratory Manual*, 2d ed., Cold

-87-

Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (see, e.g., Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (see, e.g., Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (see, e.g., Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (see, e.g., Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (see, e.g., Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (see, e.g., Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (see, e.g., Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)).

20 The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

30 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression

-88-

of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated
5 immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than
10 one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA
15 analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of
20 the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide
25 sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265:
30 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable
5 marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications
10 to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

15 In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells,
20 blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably
25 vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve
30 expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is
5 incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange
10 of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated
15 with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred
20 embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 115 and SEQ ID NO: 116 through 218 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting
25 certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and
30 "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation,

-91-

chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Breast Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a BSNA or a BSP in a human patient that has or may have breast cancer, or who is at risk of developing breast cancer, with the expression of a BSNA or a BSP in a normal human control. For purposes of the present invention, "expression of a BSNA" or "BSNA expression" means the quantity of BSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a BSP" or "BSP expression" means the amount of BSP that can be measured by any method known in the art or the level of translation of a BSG BSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing breast cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of BSNA or BSP in cells, tissues, organs or bodily fluids compared with levels of BSNA or BSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a BSNA or BSP in the patient versus the normal human control is associated with the presence of breast cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in the structure of the mRNA of a BSG compared to the mRNA from a normal

-93-

control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in a BSP compared to a BSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the BSP or subcellular BSP localization.

In a preferred embodiment, the expression of a BSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 116 through 218, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the BSNA expression that is measured is the level of expression of a BSNA mRNA selected from SEQ ID NO: 1 through 115, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. BSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. BSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a BSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, BSNA expression may be compared to a known control, such as normal breast nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a BSP is measured by determining the level of a BSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 116 through 218, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of BSNA or BSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of breast cancer. The expression level of a BSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the BSP expression level may be

determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g., Harlow (1999), supra; Ausubel (1992), supra; and Ausubel (1999), supra.* Alterations in the BSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

10 In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a BSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-BSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the
15 antibody on the solid support under conditions in which the BSP will bind to the anti-BSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-BSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the BSP to the labeled antibody will occur. After
20 binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a BSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

25 Other methods to measure BSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-BSP antibody is attached to a solid support and an allocated amount of a labeled BSP and a sample of interest are incubated with the solid support. The amount of labeled BSP detected which is attached to the solid support can be correlated to the quantity of a BSP in the sample.

30 Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size

-95-

using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with
5 chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a BSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant
10 cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is
15 then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more BSNAs of interest. In this approach, all or a portion of one or more BSNAs is fixed to a substrate. A sample of interest, which may comprise RNA,
20 *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including,
25 without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy
30 material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of BSNA or BSP includes, without limitation, breast tissue, fluid

-96-

obtained by bronchial alveolar lavage (BAL), sputum, breast cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary breast cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in BSNAs or BSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a BSNA or BSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other BSNA or BSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular BSNA or BSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having breast cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP and then ascertaining whether the patient has breast cancer from the expression level of the BSNA or BSP. In general, if high expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least

-97-

ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least
5 five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether breast cancer has metastasized in a patient. One may identify whether the breast cancer has
10 metastasized by measuring the expression levels and/or structural alterations of one or more BSNA and/or BSPs in a variety of tissues. The presence of a BSNA or BSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a BSNA or BSP is associated with breast cancer. Similarly, the presence of a BSNA or
15 BSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a BSNA or BSP is associated with breast cancer. Further, the presence of a structurally altered BSNA or BSP that is associated with breast cancer is also indicative of metastasis.

In general, if high expression relative to a control of a BSNA or BSP is indicative
20 of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is
25 considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The BSNA or BSP of this invention may be used as element in an array or a
30 multi-analyte test to recognize expression patterns associated with breast cancers or other breast related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of breast disorders.

Staging

The invention also provides a method of staging breast cancer in a human patient. The method comprises identifying a human patient having breast cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or

5 structural alterations of one or more BSNA or BSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more BSNA or BSPs is determined for each stage to obtain a standard expression level for each BSNA and BSP. Then, the BSNA or BSP expression levels are determined in a biological sample from a patient whose stage of cancer is not

10 known. The BSNA or BSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the BSNA and BSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a BSNA or BSP to determine the stage of a breast cancer.

15 *Monitoring*

Further provided is a method of monitoring breast cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also

20 monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the breast cancer. The method comprises identifying a human patient that one wants to monitor for breast cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more BSNA or BSPs, and comparing the BSNA or BSP

25 levels over time to those BSNA or BSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a BSNA or BSP that are associated with breast cancer.

If increased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then

30 detecting an increase in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a

-99-

decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of BSNA or BSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of breast cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

10 The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a BSNA and/or BSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more BSNA and/or BSPs are detected. The presence of higher (or lower) BSNA or BSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly breast cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more BSNA and/or BSPs of the invention can also be monitored by analyzing levels of expression of the BSNA and/or BSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

25 The methods of the present invention can also be used to detect genetic lesions or mutations in a BSG, thereby determining if a human with the genetic lesion is susceptible to developing breast cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing breast cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the BSGs of this invention, a chromosomal rearrangement of BSG, an aberrant modification of BSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a BSG. Methods to detect such lesions in the BSG of

-100-

this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Breast Diseases

5 The invention also provides a method for determining the expression levels and/or structural alterations of one or more BSNA and/or BSPs in a sample from a patient suspected of having or known to have a noncancerous breast disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP, comparing the
10 expression level or structural alteration of the BSNA or BSP to a normal breast control, and then ascertaining whether the patient has a noncancerous breast disease. In general, if high expression relative to a control of a BSNA or BSP is indicative of a particular noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at
15 least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of a noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even
20 more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

 One having ordinary skill in the art may determine whether a BSNA and/or BSP is associated with a particular noncancerous breast disease by obtaining breast tissue
25 from a patient having a noncancerous breast disease of interest and determining which BSNA and/or BSPs are expressed in the tissue at either a higher or a lower level than in normal breast tissue. In another embodiment, one may determine whether a BSNA or BSP exhibits structural alterations in a particular noncancerous breast disease state by obtaining breast tissue from a patient having a noncancerous breast disease of interest
30 and determining the structural alterations in one or more BSNA and/or BSPs relative to normal breast tissue.

Methods for Identifying Breast Tissue

In another aspect, the invention provides methods for identifying breast tissue. These methods are particularly useful in, *e.g.*, forensic science, breast cell differentiation
5 and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is breast tissue or has breast tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising breast tissue or having breast tissue-like characteristics, determining whether the sample expresses one or more BSNAs
10 and/or BSPs, and, if the sample expresses one or more BSNAs and/or BSPs, concluding that the sample comprises breast tissue. In a preferred embodiment, the BSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 116 through 218, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 115, or a
15 hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a BSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a BSP is expressed. Determining whether a sample
20 expresses a BSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the BSP has an amino acid sequence selected from SEQ ID NO: 116 through 218, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two BSNAs and/or BSPs is determined. In a more preferred embodiment, the
25 expression of at least three, more preferably four and even more preferably five BSNAs and/or BSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is breast tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are
30 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into breast tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new breast tissue by tissue engineering. These agents

-102-

include, e.g., growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

5 Methods for Producing and Modifying Breast Tissue

In another aspect, the invention provides methods for producing engineered breast tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a BSNA or a BSG into the cells, and growing the cells under conditions in
10 which they exhibit one or more properties of breast tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal breast tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered breast tissue or cells comprises one of these cell types. In another embodiment, the engineered breast tissue or cells comprises more than one breast cell
15 type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the breast cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more BSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules
20 encode BSPs having amino acid sequences selected from SEQ ID NO: 116 through 218, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 115, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a BSG is introduced into the cells.
25 Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial breast tissue may be used to treat patients who have lost some or all of their breast function.

Pharmaceutical Compositions

30

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a

-103-

preferred embodiment, the pharmaceutical composition comprises a BSNA or part thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 115, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a BSP or fragment thereof. In a more preferred embodiment, the BSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 116 through 218, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-BSP antibody, preferably an antibody that specifically binds to a BSP having an amino acid that is selected from the group consisting of SEQ ID NO: 116 through 218, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

-104-

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium
5 carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as
10 the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl
15 methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

20 Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

25 Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or
30 suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

-105-

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid
5 formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the
10 form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include
15 carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose
20 solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as
25 vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran.
30 Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

-106-

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

-107-

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example BSP polypeptide, fusion protein, or fragments thereof, antibodies specific for BSP, agonists, antagonists or inhibitors of BSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-

-108-

known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of breast function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g., Doronin et al., J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a BSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a BSP are administered, for example, to complement a deficiency in the native BSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes a BSP having the amino acid sequence of SEQ ID NO: 116 through 218, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a BSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in BSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a BSP having the amino acid sequence of SEQ ID NO: 116 through 218, or a fragment, fusion protein, allelic variant or homolog thereof.

-110-

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a BSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a BSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a BSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to BSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entirety.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the BSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a BSP, preferably a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a BSP, a

-111-

fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant BSP defect.

Protein compositions are administered, for example, to complement a deficiency in native BSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to BSP. The immune response can be used to modulate activity of BSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate BSP.

10 In a preferred embodiment, the polypeptide is a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of BSP, or to target therapeutic agents to sites of BSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a BSP comprising an amino acid sequence of SEQ ID NO: 116 through 218, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a BSP or have a modulatory effect on the expression or activity of a BSP.

30 Modulators which decrease the expression or activity of BSP (antagonists) are believed to be useful in treating breast cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small

-112-

molecules predicted via computer imaging to specifically bind to regions of a BSP can also be designed, synthesized and tested for use in the imaging and treatment of breast cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the BSPs identified herein. Molecules
5 identified in the library as being capable of binding to a BSP are key candidates for further evaluation for use in the treatment of breast cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a BSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of BSP is
10 administered. Antagonists of BSP can be produced using methods generally known in the art. In particular, purified BSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a BSP.

In other embodiments a pharmaceutical composition comprising an agonist of a
15 BSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP comprising an amino acid sequence of SEQ
ID NO: 116 through 218, or a fusion protein, allelic variant, homolog, analog or
20 derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 115, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Breast Tissue

25 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the breast or to specific cells in the breast. In a preferred embodiment, an anti-BSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if breast tissue needs to be
30 selectively destroyed. This would be useful for targeting and killing breast cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting breast cell function.

-113-

In another embodiment, an anti-BSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring breast function, identifying breast cancer tumors, and identifying noncancerous breast diseases.

5

EXAMPLES

Example 1: Gene Expression analysis

BSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the BSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4); differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided

30

into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into "gene bins," where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with
5 rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie "The significance of digital gene expression profiles," Genome Res 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are
10 incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were
15 analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

- (a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest
20 compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.
- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and
25 undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the
30 tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

The CLASPT™ scores for some of sequences found be the mRNA subtractions are listed below:

- DEX0267_23 Breast 5
 DEX0267_71 Breast 5
 5 DEX0267_78 Breast 5 and 1
 DEX0267_89 Breast 5
 DEX0267_101 Breast 5

The CLASPT™ expression levels for selected sequences are listed below:

- | | | | | | | |
|----|-------------|----------------|------------|-----------|-----------|-----------|
| | DEX0267_11 | SEQ ID NO: 11 | BRN .0002 | LNG .0011 | FAL .0063 | ESO .0102 |
| 10 | DEX0267_23 | SEQ ID NO: 23 | MAM .0179 | TST .0011 | BLO .0019 | SPL .002 |
| | DEX0267_61 | SEQ ID NO: 61 | MAM 1.0726 | NOS .3813 | PLE .4337 | PIB .5075 |
| | DEX0267_66 | SEQ ID NO: 66 | BRN .0002 | LNG .0011 | PRO .0011 | |
| | DEX0267_67 | SEQ ID NO: 67 | LMN .0028 | URE .0112 | UNC .016 | |
| | DEX0267_71 | SEQ ID NO: 71 | MAM .0142 | UTR .0094 | ADR .0179 | |
| 15 | DEX0267_73 | SEQ ID NO: 73 | MAM .0028 | UTR .0006 | THY .002 | OVR .0031 |
| | DEX0267_76 | SEQ ID NO: 76 | INS .0076 | | | |
| | DEX0267_78 | SEQ ID NO: 78 | MAM .0014 | FTS .0001 | UTR .0004 | PRO .0007 |
| | DEX0267_80 | SEQ ID NO: 80 | UTR .0006 | BLD .0048 | FAL .0063 | CRD .0068 |
| | DEX0267_89 | SEQ ID NO: 89 | MAM .0094 | SPL .0063 | OVR .0092 | PNS .0094 |
| 20 | DEX0267_93 | SEQ ID NO: 93 | TST .0054 | | | |
| | DEX0267_94 | SEQ ID NO: 94 | TST .0054 | | | |
| | DEX0267_98 | SEQ ID NO: 98 | MAM .3287 | SAG .079 | UNC .1635 | PIT .2054 |
| | DEX0267_100 | SEQ ID NO: 100 | PNS .0164 | LMN .0222 | OVR .0246 | NOS .0587 |
| | DEX0267_101 | SEQ ID NO: 101 | MAM .0061 | STO .0081 | FAL .0126 | URE .0337 |
| 25 | DEX0267_115 | SEQ ID NO: 115 | MAM .0128 | ADR .0015 | LIV .0019 | SPL .0021 |

Abbreviation for tissues:

- ADR Adrenal Glands, BLD Bladder, BLO Blood, BRN Brain, CRD Heart, CTL
 Cartilage, ESO Esophagus, FAL Fallopian Tubes, FTS Fetus, GEM Germ Cells, INS
 Intestine, Small, INT Intestine, LIV Liver, LMN Lymphoid Tissue, LNG Lung, MAM
 30 Breast, NOS Nose, OVR Ovary, PIB Pineal Body, PIT Pituitary Gland, PLE Pleura, PNS
 Penis, PRO Prostate, SAG Salivary Glands, SPL Spleen, STO Stomach, THY Thymus
 Gland, TST Testis, UNC Mixed Tissues, URE Ureter, UTR Uterus

The chromosomal locations for the sequences are as follows:

- 35 DEX0267_2 chromosome 2
 DEX0267_12 chromosome 9
 DEX0267_23 chromosome 4
 DEX0267_31 chromosome 10
 DEX0267_36 chromosome 16
 40 DEX0267_44 chromosome 10
 DEX0267_72 chromosome 15
 DEX0267_73 chromosome 1
 DEX0267_94 chromosome 2
 DEX0267_96 chromosome 14
 45 DEX0267_103 chromosome 16

-116-

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the BSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to a normal tissue (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the BSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to a normal tissue (calibrator). A matching pair is formed by mRNA from

-117-

the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, BSNA's show a high degree of tissue specificity for the tissue of interest. Results from these experiments confirm the tissue
5 specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (e.g. higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

10 Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 115 being a diagnostic marker for cancer.

Example 3: Protein Expression

The BSNA is amplified by polymerase chain reaction (PCR) and the amplified
15 DNA fragment encoding the BSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the BSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of BSNA, and six histidines, flanking the COOH-terminus of the coding sequence of BSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

20 An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of BSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography
25 (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickel chelating resin. The column was packed and washed with five column volumes of wash buffer. BSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

30 Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an

-118-

expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a

5 polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the

10 vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such

15 cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any

20 suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

25 The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to

30 obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the

-119-

hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

- 5 Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

	DEX0267_116	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
10	18-28	1.01	11	
	DEX0267_118	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	12-29	1.01	18	
	DEX0267_120	Antigenicity Index(Jameson-Wolf)		
15	positions	AI avg	length	
	150-162	1.30	13	
	55-65	1.09	11	
	3-51	1.03	49	
	101-123	1.03	23	
20	DEX0267_122	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	23-32	1.05	10	
	DEX0267_125	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
25	221-233	1.16	13	
	124-142	1.16	19	
	279-289	1.14	11	
	261-271	1.10	11	
	DEX0267_129	Antigenicity Index(Jameson-Wolf)		
30	positions	AI avg	length	
	7-48	1.13	42	
	DEX0267_133	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
35	398-409	1.30	12	
	22-38	1.21	17	
	478-489	1.15	12	
	90-103	1.10	14	
	111-134	1.06	24	
	376-396	1.05	21	
40	319-328	1.04	10	
	331-366	1.02	36	
	DEX0267_138	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	67-77	1.01	11	
45	DEX0267_140	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	30-42	1.17	13	
	DEX0267_141	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
50	100-115	1.10	16	
	DEX0267_143	Antigenicity Index(Jameson-Wolf)		
	positions	AI avg	length	
	108-118	1.10	11	

-120-

	166-216	1.02	51
	DEX0267_144	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	17-26	1.06	10
5	DEX0267_146	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	8-58	1.06	51
	DEX0267_148	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
10	41-56	1.15	16
	DEX0267_153	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	39-73	1.13	35
	DEX0267_155	Antigenicity Index(Jameson-Wolf)	
15	positions	AI avg	length
	7-32	1.11	26
	56-71	1.00	16
	DEX0267_156	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
20	7-19	1.06	13
	DEX0267_158	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	98-118	1.00	21
	DEX0267_167	Antigenicity Index(Jameson-Wolf)	
25	positions	AI avg	length
	17-28	1.14	12
	DEX0267_170	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	55-68	1.36	14
30	18-43	1.12	26
	DEX0267_171	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	88-107	1.16	20
	DEX0267_175	Antigenicity Index(Jameson-Wolf)	
35	positions	AI avg	length
	108-119	1.10	12
	DEX0267_179	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	358-388	1.20	31
40	311-342	1.11	32
	218-230	1.05	13
	18-37	1.00	20
	DEX0267_182	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
45	162-176	1.11	15
	DEX0267_191	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	5-33	1.12	29
	DEX0267_192	Antigenicity Index(Jameson-Wolf)	
50	positions	AI avg	length
	187-207	1.11	21
	44-56	1.09	13
	DEX0267_194	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
55	46-61	1.15	16
	74-96	1.13	23
	DEX0267_196	Antigenicity Index(Jameson-Wolf)	
	positions	AI avg	length
	8-29	1.16	22

-121-

	DEX0267_197	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	26-35	1.06 10
	90-101	1.05 12
5	DEX0267_199	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	5-25	1.14 21
	27-42	1.10 16
10	DEX0267_201	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	123-138	1.15 16
	DEX0267_202	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	15-32	1.25 18
15	DEX0267_205	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	14-23	1.03 10
	DEX0267_206	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
20	8-23	1.19 16
	DEX0267_208	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	30-39	1.23 10
	11-27	1.07 17
25	DEX0267_210	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	56-67	1.17 12
	DEX0267_211	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
30	35-44	1.05 10
	DEX0267_212	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	80-89	1.12 10
	43-68	1.07 26
35	95-108	1.04 14
	DEX0267_213	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	114-123	1.33 10
	DEX0267_214	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
40	22-36	.15 15
	DEX0267_215	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	17-27	1.00 11
45	DEX0267_218	Antigenicity Index(Jameson-Wolf)
	positions	AI avg length
	26-46	1.10 21

Examples of post-translational modifications (PTMs) of the BSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the LSPs of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html)

most recently accessed October 23, 2001). For full definitions of the PTMs see

<http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

- DEX0267_117 Camp_Phospho_Site 10-13;
- DEX0267_118 Ck2_Phospho_Site 45-48; Myristyl 27-32;32-37; Pkc_Phospho_Site 13-15;99-101;
- 5 DEX0267_119 Ck2_Phospho_Site 32-35; Myristyl 49-54;
- DEX0267_120 Amidation 86-89; Asn_Glycosylation 90-93; Camp_Phospho_Site 105-108;
Ck2_Phospho_Site 125-128;174-177; Myristyl 71-76;159-164;184-189;
Pkc_Phospho_Site 103-105;
- DEX0267_121 Asn_Glycosylation 27-30; Ck2_Phospho_Site 29-32; Pkc_Phospho_Site 14-16;
- 10 DEX0267_122 Camp_Phospho_Site 73-76; Ck2_Phospho_Site 23-26;102-105; Myristyl 4-9;55-60;84-
89; Pkc_Phospho_Site 23-25;69-71;88-90;113-115;
- DEX0267_124 Asn_Glycosylation 36-39;
- DEX0267_125 Asn_Glycosylation 56-59;268-271;283-286; Camp_Phospho_Site 191-194;221-224;
Ck2_Phospho_Site 106-109;136-139;147-150;255-258; Glycosaminoglycan 231-234;
Pkc_Phospho_Site 66-68;69-71;147-149;
- 15 DEX0267_126 Myristyl 16-21;55-60;
- DEX0267_127 Asn_Glycosylation 23-26;
- DEX0267_128 Cytochrome_C 36-41; Myristyl 2-7;4-9;63-68;
- DEX0267_129 Camp_Phospho_Site 9-12; Ck2_Phospho_Site 60-63;76-79; Myristyl 28-33;
- 20 DEX0267_130 Pkc_Phospho_Site 7-9;12-14;46-48;
- DEX0267_131 Myristyl 16-21;38-43;
- DEX0267_131 Amidation 19-22; Ck2_Phospho_Site 76-79; Myristyl 36-41;37-42; Pkc_Phospho_Site
13-15;76-78;
- DEX0267_132 Myristyl 15-20;
- 25 DEX0267_133 Asn_Glycosylation 98-101;289-292;322-325; Ck2_Phospho_Site 2-5;80-83;199-
202;217-220; Myristyl 8-13;41-46;97-102;187-192;251-256;252-257;287-292;484-489;
Pkc_Phospho_Site 28-30;29-31;34-36;110-112;113-115;124-126;199-201;239-241;296-
298;327-329;
- DEX0267_134 Myristyl 53-58;
- 30 DEX0267_135 Myristyl 61-66;
- DEX0267_136 Asn_Glycosylation 65-68; Camp_Phospho_Site 20-23;26-29; Myristyl 46-51;
Pkc_Phospho_Site 23-25;
- DEX0267_137 Asn_Glycosylation 82-85;85-88; Ck2_Phospho_Site 15-18;33-36;48-51; Myristyl 27-32;
Pkc_Phospho_Site 15-17;23-25;57-59;81-83;
- 35 DEX0267_138 Myristyl 38-43;
- DEX0267_139 Ck2_Phospho_Site 7-10;
- DEX0267_140 Myristyl 13-18;27-32;
- DEX0267_141 Camp_Phospho_Site 78-81; Pkc_Phospho_Site 99-101;105-107;
- DEX0267_142 Myristyl 24-29; Pkc_Phospho_Site 17-19;49-51;
- 40 DEX0267_143 Amidation 64-67;149-152; Camp_Phospho_Site 99-102;181-184; Myristyl 42-47;45-
50;212-217;213-218; Pkc_Phospho_Site 14-16;97-99;112-114;131-133;132-134;159-
161;
- DEX0267_144 Ck2_Phospho_Site 3-6; Pkc_Phospho_Site 3-5;9-11;
- DEX0267_145 Amidation 178-181; Ck2_Phospho_Site 274-277; Myristyl 39-44;102-107;174-179;197-
202; Pkc_Phospho_Site 215-217;247-249;278-280; Prokar_Lipoprotein 30-40; Rgd 166-
168;183-185;
- 45 DEX0267_146 Ck2_Phospho_Site 16-19;86-89; Pkc_Phospho_Site 79-81;92-94;
- DEX0267_147 Ck2_Phospho_Site 36-39; Myristyl 72-77; Pkc_Phospho_Site 29-31;42-44;45-47;
- DEX0267_148 Asn_Glycosylation 13-16; Camp_Phospho_Site 28-31; Ck2_Phospho_Site 75-78;
- 50 DEX0267_149 Ck2_Phospho_Site 3-6; Myristyl 9-14; Pkc_Phospho_Site 27-29;
- DEX0267_150 Ck2_Phospho_Site 9-12;21-24; Pkc_Phospho_Site 18-20;28-30;34-36;
- DEX0267_151 Myristyl 22-27;
- DEX0267_152 Glycosaminoglycan 3-6;9-12;
- DEX0267_153 Amidation 67-70; Camp_Phospho_Site 69-72; Myristyl 64-69; Pkc_Phospho_Site 30-
32;56-58;
- 55 DEX0267_154 Asn_Glycosylation 12-15; Myristyl 51-56;

-123-

- DEX0267_155 Asn_Glycosylation 65-68; Ck2_Phospho_Site 24-27;50-53; Myristyl 98-103;
Pkc_Phospho_Site 57-59;70-72;
DEX0267_156 Pkc_Phospho_Site 10-12;64-66;
DEX0267_157 Asn_Glycosylation 27-30;
5 DEX0267_158 Ck2_Phospho_Site 125-128; Pkc_Phospho_Site 32-34;77-79;125-127;
DEX0267_159 Ck2_Phospho_Site 53-56;97-100; Pkc_Phospho_Site 93-95;
DEX0267_160 Ck2_Phospho_Site 5-8;
DEX0267_162 Amidation 19-22; Camp_Phospho_Site 22-25; Myristyl 9-14; Rgd 79-81;
DEX0267_163 Ck2_Phospho_Site 37-40;
10 DEX0267_165 Myristyl 24-29;
DEX0267_166 Myristyl 17-22;
DEX0267_167 Ck2_Phospho_Site 59-62;
DEX0267_168 Asn_Glycosylation 64-67; Myristyl 62-67; Tyr_Phospho_Site 47-54;
DEX0267_169 Amidation 179-182; Camp_Phospho_Site 11-14;68-71;69-72;189-192;
15 DEX0267_170 Ck2_Phospho_Site 42-45;80-83;116-119;124-127; Myristyl 144-149; Pkc_Phospho_Site
7-9;17-19;42-44;65-67;72-74;80-82;116-118;124-126;157-159;187-189;192-194;203-
205; Rgd 38-40;183-185;
DEX0267_170 Asn_Glycosylation 50-53; Pkc_Phospho_Site 28-30;
DEX0267_171 Ck2_Phospho_Site 2-5;120-123;140-143; Myristyl 73-78;79-84;110-115;
20 DEX0267_172 Pkc_Phospho_Site 8-10;19-21;39-41;92-94;120-122;
DEX0267_172 Myristyl 5-10;
DEX0267_173 Ck2_Phospho_Site 40-43; Pkc_Phospho_Site 13-15;
DEX0267_175 Ck2_Phospho_Site 4-7; Myristyl 115-120;121-126; Pkc_Phospho_Site 93-95;
DEX0267_176 Myristyl 108-113;
25 DEX0267_178 Amidation 67-70;94-97;122-125; Camp_Phospho_Site 32-35;57-60;75-78;103-106;114-
117;119-122;175-178; Ck2_Phospho_Site 2-5;60-63;82-85;86-89;132-135;143-146;155-
158;183-186;195-198;204-207; Pkc_Phospho_Site 26-28;31-33;37-39;41-43;56-58;86-
88;106-108;117-119;122-124;132-134;143-145;178-180;194-196;195-197;
Tyr_Phospho_Site 142-149;
30 DEX0267_179 Asn_Glycosylation 393-396; Camp_Phospho_Site 406-409; Ck2_Phospho_Site 46-
49;143-146;164-167;238-241;312-315;362-365;384-387; Glycosaminoglycan 214-217;
Myristyl 52-57;156-161;160-165;274-279; Pkc_Phospho_Site 157-159;208-210;222-
224;349-351;408-410;409-411;418-420;
DEX0267_180 Ck2_Phospho_Site 36-39;
35 DEX0267_181 Ck2_Phospho_Site 46-49;
DEX0267_182 Asn_Glycosylation 172-175; Ck2_Phospho_Site 141-144;170-173; Myristyl 176-181;
Pkc_Phospho_Site 29-31;67-69;141-143; Prokar_Lipoprotein 110-120;
DEX0267_184 Ck2_Phospho_Site 22-25; Myristyl 99-104;
DEX0267_185 Ck2_Phospho_Site 21-24;
40 DEX0267_186 Asn_Glycosylation 17-20; Pkc_Phospho_Site 31-33;41-43;50-52;
DEX0267_189 Myristyl 6-11;
DEX0267_190 Camp_Phospho_Site 62-65;63-66; Myristyl 13-18; Pkc_Phospho_Site 14-16;66-68;72-
74;
DEX0267_191 Asn_Glycosylation 11-14;34-37; Pkc_Phospho_Site 17-19;36-38;
45 DEX0267_192 Ck2_Phospho_Site 24-27;148-151;231-234;257-260; Glycosaminoglycan 4-7; Myristyl
5-10;79-84;144-149;149-154;184-189; Pkc_Phospho_Site 9-11;
DEX0267_193 Myristyl 22-27;26-31; Prokar_Lipoprotein 42-52;84-94; Receptor_Cytokines_1 45-57;
DEX0267_194 Ck2_Phospho_Site 35-38; Myristyl 7-12;28-33;50-55;61-66; Pkc_Phospho_Site 31-
33;51-53;65-67;126-128;
50 DEX0267_195 Ck2_Phospho_Site 35-38; Myristyl 31-36;74-79;
DEX0267_196 Camp_Phospho_Site 20-23; Ck2_Phospho_Site 23-26; Myristyl 29-34;
DEX0267_197 Asn_Glycosylation 93-96;94-97; Ck2_Phospho_Site 9-12;89-92;162-165;229-232;
Pkc_Phospho_Site 72-74;124-126;143-145;
DEX0267_199 Camp_Phospho_Site 34-37; Myristyl 6-11; Pkc_Phospho_Site 18-20;37-39;
55 DEX0267_200 Pkc_Phospho_Site 21-23;112-114; Prokar_Lipoprotein 230-240;
DEX0267_201 Amidation 124-127; Ck2_Phospho_Site 68-71; Pkc_Phospho_Site 137-139;
DEX0267_202 Asn_Glycosylation 53-56; Myristyl 30-35; Pkc_Phospho_Site 3-5;15-17;
DEX0267_204 Ck2_Phospho_Site 56-59;
DEX0267_205 Ck2_Phospho_Site 29-32;

-124-

- DEX0267_206 Ck2_Phospho_Site 16-19;23-26; Myristyl 21-26;
 DEX0267_207 Asn_Glycosylation 8-11; Ck2_Phospho_Site 13-16;31-34; Myristyl 19-24;
 DEX0267_208 Amidation 34-37; Myristyl 8-13;9-14;61-66; Pkc_Phospho_Site 45-47;53-55;
 DEX0267_209 Myristyl 25-30;35-40;39-44; Pkc_Phospho_Site 13-15;57-59;
 5 DEX0267_210 Asn_Glycosylation 26-29; Pkc_Phospho_Site 15-17;46-48;65-67; Tyr_Phospho_Site 73-80;
 DEX0267_211 Ck2_Phospho_Site 6-9;58-61; Glycosaminoglycan 92-95; Myristyl 15-20;59-64;86-91;
 Pkc_Phospho_Site 120-122; Tyr_Phospho_Site 111-119;
 10 DEX0267_212 Camp_Phospho_Site 58-61;113-116; Myristyl 100-105; Pkc_Phospho_Site 61-63;97-99;107-109;116-118;
 DEX0267_213 Camp_Phospho_Site 115-118; Myristyl 126-131; Pkc_Phospho_Site 40-42;114-116;118-120; Tyr_Phospho_Site 81-88;
 DEX0267_214 Amidation 27-30; Ck2_Phospho_Site 5-8;76-79;111-114; Myristyl 70-75;
 Pkc_Phospho_Site 23-25;85-87;111-113;
 15 DEX0267_215 Ck2_Phospho_Site 54-57; Pkc_Phospho_Site 25-27;
 DEX0267_217 Camp_Phospho_Site 87-90; Ck2_Phospho_Site 27-30;104-107;105-108; Myristyl 5-10;9-14; Pkc_Phospho_Site 26-28;101-103;104-106;

Example 6: Method of Determining Alterations in a Gene Corresponding to a

20 Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1
 25 through 115. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4
 30 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156
 35 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99

-125-

(1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping
5 are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the
10 genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

15 Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The
20 coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature.
25 The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is
30 prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear

scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 $\mu\text{g/kg/day}$ to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 $\mu\text{g/kg/hour}$ to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers

-127-

of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-
5 release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and
10 EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is
15 formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and
20 other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such
25 carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at
30 the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic

-128-

polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a

pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 $\mu\text{g/kg}$ of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

-130-

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

-131-

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al. (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) *Ann. NY Acad. Sci.* 772: 126-139 and Abdallah B. et al. (1995) *Biol. Cell* 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or

-132-

chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They
5 may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and
10 express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of
15 ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other
20 parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is
25 determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

30 Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute,

-133-

approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared
5 by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and
10 HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals.
15 Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.
20 Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-
25 834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a
30 gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl.

Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated
5 oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene
10 or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of
15 interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal
20 sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell
25 type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression
30 of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated

-135-

immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive

-136-

targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*).

However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

5 In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts,
10 bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral
15 vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve
20 expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

 Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft;
25 genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

 When the cells to be administered are non-autologous or non-MHC compatible
30 cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange

This page is not part of
the pamphlet!

WO 02-068645

2/4

Date: 06 sep 2002

Destination: Agent

Address:

-137-

of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

15

-138-

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 116 through 218;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 115;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
 - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a breast specific nucleic acid
(BSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to a breast
specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to a BSNA in the
sample, wherein the detection of the hybridization indicates the presence of a BSNA in
the sample.

-139-

7. A vector comprising the nucleic acid molecule of claim 1.
8. A host cell comprising the vector according to claim 7.
- 5 9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.
- 10 10. A polypeptide encoded by the nucleic acid molecule according to claim 1.
 11. An isolated polypeptide selected from the group consisting of:
 - (a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 116 through 218; or
 - 15 (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 115.
 12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.
- 20 13. A method for determining the presence of a breast specific protein in a sample, comprising the steps of:
 - (a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the breast specific protein; and
 - 25 (b) detecting binding of the antibody to a breast specific protein in the sample, wherein the detection of binding indicates the presence of a breast specific protein in the sample.
- 30 14. A method for diagnosing and monitoring the presence and metastases of breast cancer in a patient, comprising the steps of:
 - (a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

-140-

- (b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the breast specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of breast cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with breast cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the breast cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> Salceda, Susana
Macina, Roberto
Recipon, Herve
Cafferkey, Robert
Sun, Yongming
Liu, Chenghua
Turner, Leah
diaDexus, Inc.

<120> Compositions and Methods Relating to Breast Specific Genes

<130> DEX-0268

<150> 60/249,992

<151> 2000-11-20

<160> 218

<170> PatentIn version 3.1

<210> 1

<211> 1767

<212> DNA

<213> Homo sapien

<400> 1

```
cggccgccccg ggcaggtaca agctttttttt tttttttttt ttttttttta aaaaactaaa      60
gtcaaattttt tttttttccc ataaaaccgc ttctctttttt attaataaaa aaaataaaaa      120
taaaaagtgg aaccaaagag gaaaaggggt ggttttaaga ggtggaccgc tgggtgggaaa      180
gagagagggcg agagggcggtg cgaggacacg agaaagaaca cgcgtgggaa cacgtgggag      240
gtggcccccg gggacacctc gagagagagg cagagagtgg cgtgtattca cacgtctctca      300
tcatgagtgg tgacacaccg agactcgcgt ggcgccgcgc ggcgtgtgtg tctcccagag      360
agagagagag ggcgtgtgta agatcatcac gcggtgggac actctcagca ggggcggtgt      420
gatgacgcc agtgtgtcgc actctgtgtg ccaccgctgt gtgtgagtgt gagagagggc      480
gactattctc ttatagagca gagagacacc ctgtgtgaga ctgtgtggga gaaaaagtgt      540
gtcgcgccac cacacacaac tctcccgcca gaggctctct gtgtgtgaga gaggagagta      600
gtatataaga ggagggacag cggcgggggg tgtatataaa ttttatctca catatttata      660
agccggtgtg tgggtgtgatg tgagagggga ggggagagag tgtcatcttc tctcacacag      720
cggagagaga gagacgggtgt gtgagggacg gcgtgtggta gtttttcttc tcctcgccgc      780
cgaagaagaa gatgttataa caaaagaagt tgtggggggc gcgcacacca aaataataga      840
aggattgttg tcgtgtgaga taatcctcga ccgcagaggc gcgcctctgc tcttctctta      900
ttatgaggtg ctacgattaa tacccccac gattgtgttt atataatcac gccgactgtt      960
```

```

gctgtctccc gacgaagggg acgggcgaag ctgctccaa tgggtggggg cccccacaaa 1020
gaggagcaac aaagaggaga acgacgtggt agcagcacgt cataataaag acgggttgta 1080
ctaacgaggg ggggaaaaca actgctggtg tggaacacgg cggggggggg ggggggtggg 1140
tcgcaccccc caaaataatt aacaccgcca gaacgaagaa gctctcacgc atcatccgct 1200
gcgaaaacac gcggccttct gtgggcgtac ttagatgcag gcgggcgtgg tttttctccc 1260
ccacgaagtg gtgatgtgtg ctcccccccg aggggggagg gagtaattat aaacaccccc 1320
ctctctgtgg gggtgagaac acaataatt gtctgctgta ggggtgggtg acaccacat 1380
cgtcagcaag agatctgtcc tggtgtgctg acaaccagc gtgtgtgtgg ggcgggcccc 1440
cctacaagag gatcagctcg cggtgtcgtt ggtataataa acaaccacac cgggggcgca 1500
gcgaggagga aaaacaaccc gtgcaggggc gtgctggcag aacaacagca gcggggaaga 1560
agattgcacc acgagtggga caaagacgga caggagcgt cgcacggcaa aatcttgctg 1620
gggcgggaaa caacaaaaca gctgcgagca gcggctggct gcgggcgtcc acaaacgatg 1680
cgtgtgcggg tgccctctc ccccagagg tcgggggcgg cggcaacaca cagggagggc 1740
aaacaacgag cgagagtcac cgggtga 1767

```

```

<210> 2
<211> 541
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (495)..(495)
<223> a, c, g or t

```

```

<400> 2
gcgtggtcgc ggcgaggtag agtcagatc ttttctttaa ttcttatggt tttttttttt 60
tttttttttt ttaaaaaatg gagtttgtgc aattttgcc aagttgattt tgaattccgg 120
ggcccaattg atccccccac ctacgcctcc tgagtggtag ggtttacggg ggtaaccat 180
tgtgcctggt ttccagcttt ccttttaaat tagggggtta tagttcgga caaccaggac 240
ccagggcagg aaatatacac ttcccccaata gcaaattagc attaccgtga cctcctctgt 300
gctaatatgg cacttttggt aaccaagtga attgatgggt gtggagtgg gtggatgtag 360
atgaagtga ttgaacata tactacgtga taatttatat ccagagtc tcaaaaatat 420
tggtggcgtt gaaaaattgg ggagggcggg agtggaatt cactgttga tatagattaa 480
ccacggtgaa attantggct tgcttgaaaa ggtcttaaag taagtgggtg tttttactca 540

```

g

541

<210> 3
 <211> 874
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (770)..(770)
 <223> a, c, g or t

<400> 3
 ctgagtttca gtcctggctc tacccttctt ggctgtggt tctaagcatg ttatttgcct 60
 ctctcagctt caactgtgaa gagttcaatt aggtgatcac tttaactttt ctagctcgga 120
 tactctgtgc cagctctgga accatgcttt ttggtgtctg tgtgtatata taggtcacct 180
 gtatgtattht aggtcctttg aggaatctac tggacgtatc aaaaaaaaaa aaaaaacacc 240
 cacaaaaaga acagccccgt ggagctcttg agtgtgggtc tccacttagt gttgtgttgt 300
 gtttctcccc aatctctttc ttagaagcca gggaggggca cccttctgtg gggctctcca 360
 ccattcttct tgaggcgagc cattccccag ccttccttct tcttccaag cctgtgttct 420
 tgttacactt ggggtgaagg gggaagtgtg ttcccggtc ggagaactgg tgtttaacag 480
 gtaaggctct tggccctccc aggtgactct ttttagggg caggaccca ttcttggtaa 540
 gcccagcatt ggctctggcc ccagacactt tgtggtttgg tctcaggtaa tcggtggctg 600
 tccactaggc tgcttggttg acctttcttg cgtggtgtcc atattggtct tcctttgtgc 660
 ggaaaattaa ttcctttcgc acttgccaca aaaaaccaa aacacaaaaa aggcgtgggg 720
 cgcccggtgc ctaagcgggt ccgtgggaga aatggttccg cccacaacn accgccacac 780
 accacacaca gcgcgggcgg gggggcgctt aaaacagaac gaagggggac gacaggcaca 840
 caaggcagga ggaacagaga aaaaggggag agtg 874

<210> 4
 <211> 557
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (404)..(486)
 <223> a, c, g or t

<220>

<221> misc_feature
 <222> (489)..(489)
 <223> a, c, g or t

<400> 4
 gcatataatg tatattgttg aatccaatca ggatgcatgt gagatgatat attggctgaa 60
 gctaccathtt taccgctgtg gctccctgag actcttgatt ctagcttctg tgtctgcgaa 120
 cgtgataact ggaggaatac tatcatagga atggtatata cgcattatga ggcacaaagt 180
 tggagtgaat gaaagcgtac tgattggagt tagaccagta gcactgaaca tagtgagtgc 240
 acgagtacat ctatacccca acaaatagtc gatcactaca tcctggaagc ataccagcac 300
 ccaagcaaca acaagacatt aggctactag caatgggggtt atattacaat taccttgact 360
 agacacataa aagaacaatt tcagagccca catgatttta gtannnnnnn nnnnnnnnnn 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
 nnnnnnatnc cccacccac caccacccc caacccccct cccaccccc ccaaccccca 540
 atagccccc cccacaa 557

<210> 5
 <211> 504
 <212> DNA
 <213> Homo sapien

<400> 5
 atgtctatgg ggactggtgt ctcctagatg ctgctcagcg gccgccaggt tgtgatggat 60
 gcgtgtcgcg gccgaggtac ctgagatatg gtcagattct aaatacathtt tagtggcaca 120
 atctacaaaa cttaatgact caccagacat gggaaatgaa agaagcagag tcctgagata 180
 acctaaagtt cttggcctga gcagctggaa gactggagtg gccatttact gagacagaga 240
 agctatgaga agaaccathtt tgggggagaa gagaacatac tgcgttgag aagtctatta 300
 gatccggttg aagatgttga gtagctathtt ggatatgtag cttttctcac agttccccaa 360
 aactttacga tttgcctacc gactgagcca acagctaaat gtgtgccctg tttttaattc 420
 tatgtgtagt ttgctgtaga aagagaaagc aactcttaaa acctgaaaag aaatgaaaat 480
 ggaaactaat gaattacatt aaga 504

<210> 6
 <211> 795
 <212> DNA
 <213> Homo sapien

<400> 6
 tttttttttt tttatttttt tttgggaaaa aacagaaccc ccccaaaaac attaatTTTT 60


```

tttttaaaaa aatctgggccc cccttgggtg gtttccaatt tggtttcccc ccttttcccc 120
ttgaacccaa attcctaaaa cttgtttttc ttaaaaaatg agttgtggct acctttaacc 180
cataccctta actcgggtgg tgtcccatat agttgctccc accccagtag ccagctctct 240
cctccaccct ttctctgcgg gtttccagtg ctctcaggg ccgtgagcag cacgtgaggg 300
gctgggacga ttttttctcc tttaacgaat gtccagctct ccagccaagt ttggagagcc 360
ttctctctca tgtgttggct caaacgcaac cgagaacgag tgtttctctt gcgggttcga 420
taggccaata ggaggaccag attgttggat ttattttttc tcccagtagt attatcccgg 480
actctaacca ctggtccaga ggcttgggtg cccaccaaca ctatatatct cctcgggggg 540
accttttcgt gggccctatc gagggagAAC agcgggtgtc tgtgcccagt tggttcctt 600
aaacccccgt atggggggga gggacaaaac gtggtctcct cgcccaattt tcggggatgt 660
ctccttttat tttccagcaa ccactttttc ttcaaaaagc tgggggggta acctggggcc 720
ataggcctgg tccccgtgg tgtaatttgg tcttccggtt ccaatttccc ccctactcac 780
agcacacccc accta 795

```

<210> 7
 <211> 260
 <212> DNA
 <213> Homo sapien

```

<400> 7
gccgggcagg taccttatat tagttttctt atttattttc acagcatcct ttttctatgt 60
agcaatgagt tgcttttttt ttgccttttt aaagatggaa gtcacagcaa aatgggaaat 120
taacttgcct attaattcat gcaacatgac aactgcagag caatgtctag agtaagacaa 180
tagtatgtct tattcttctt cagaaaatat tcttatatgt catatttagt taaaatatca 240
tgtatcatat catatgttta 260

```

<210> 8
 <211> 609
 <212> DNA
 <213> Homo sapien

```

<400> 8
gcgatgttca tcaactatag gcgaatggc cctagatgca tgccgagcgg cgcaggttgt 60
gatggatcgg cgcccgggca ggtacattgt tttttttttt tttttttttt ttttttgaaa 120
aaaacccccg ttttaatacc ttattttttt tggttttaaa aaaatttttt aaccatttta 180
aaaaaacccc ccctttcccc catttcagtt tccccgttaa acgggtttta aagttgaggg 240

```

6

```

aaagtgaatt tttgtctcca ccgagctttg ggaccactca gcggttccgt gtgcaaagga    300
ccttctcgag acaccaaccc cctttgtgcc aaaaaaatc gtggacagct ttttacctt    360
gttggtctta taaacaaata ccagacgcgg gatattctcc cccccccctc gtagatgtgg    420
gacaaacccg ccttgtctca ccagccaaat ctttctctct ccacccaaac acgagagctg    480
tggggggtat acatctcgag tggctctcaa tagcgtctg ttccacgcgt ggtgtgtaga    540
aatgtgtgtt tctctcgcg cttctcaacat atctcccacc aaaaaattag cacaacacaa    600
aatggaatg                                     609

```

<210> 9
 <211> 450
 <212> DNA
 <213> Homo sapien

```

<400> 9
actaatcatt atttctttct tttttttttt tttggggagg gagctcttgc tctgtcacc    60
aggcgggaat tgtcgggggt gcaatcttgg gctcacgtgg aacctcctcc tcttgggggt    120
caaggtgatt ctccgtggtg cctcagccct cccgagttgg ggggccccg ggtgcccggt    180
accagtgccc gggttaattt ctgggtatat ttaaggtaga agaacgaggg ttctcaccat    240
tgttgggcca ggcgggtctc aaactcctg gacttcaagt gatctgcca tctgggactc    300
ccaaagggcg gtgggattac gaggttgag ccaccatag cggccgattt tataatgata    360
ctctaaataa cacttttctt acactgggat ttgcccaaag atcattgggt gaaccttcc    420
cacccttggt tttgtgaagc aaacggaact                                     450

```

<210> 10
 <211> 238
 <212> DNA
 <213> Homo sapien

```

<400> 10
atccttatta gatatgtaat ttggcacata ttttctcca ttttgtgggt tgtctttgtc    60
ttttttattt tttattgctt gtgtttggt tctatttga taattcattg ttaaattaaa    120
ggtcatgaac gagttaccaa aaaaacacaa cagaaaaaaaa aaaaagcctg ggggaaaacc    180
aggccaaacc gtttccccg ggggaaattg tttccgccac attcaataaa acaaaaaac    238

```

<210> 11
 <211> 1925
 <212> DNA
 <213> Homo sapien

<400> 11

tttttttttg aatgtttata acagctttat taatattggc caaaacttgg aagcaaccaa 60
gatgtccctc tataggtgca tagataaaca ttttatggcc catccataaa atgaaacatt 120
attcagcaat aaaaggaaat gaggtataaa gccatgaaga gatatggggg aaattttaa 180
tcatattgct aagtgagaga agccagtttg ttagtttatt ttataaatca ggatatgggt 240
tattttgggtg aatattccat gtgtacttgc aaaaatttg aattctacca cttttgggtta 300
tagtgggtcta taaatgtcca ttaggacaag tttatcctag tgttggtcag atctatcctt 360
gttaactttt tagctaattt atttagctaa aattaatttt ttagctaact tttattaatt 420
attaagagta aagcatttaa atcccaaata taattgtgga tttgtcaatt tccccttgca 480
gttatgtcaa tttttacttc cggtattgtt atgtctttat cagggggcaa ttcaccaagc 540
agtgcctctt ttctctctga taatattgct tcttttgaag tccacttttg tttatattaa 600
tattgtcatt ccaacttttt tttgactaga gtttgcacaa tatattatcc ctttcttttt 660
atttacagtg agttacgggt agaaagcata tatttgggtc tctaggaaga attgaatatt 720
taatacgtgg gacaatataa gatttctatt ttttcttgag taggttttga taatttatat 780
tttctaggaa tttgtctatt ttctctaaac tttcaaacc ctttggcata aattgttaac 840
actgtccctt aatcttttta atctttatgg tgttttcaa tatgtctccc ctttcttttc 900
ataatattat ttctatatac ttttcttttt gtcttgatta atcggccaaa tgtttgtcta 960
ttttattaca aaactaaaat aaccaaaca ggctgggtact ggcataaaga taaacattta 1020
gaccaataga atagaattga gactgcagaa gtaaacatcat acatatatct tcaattgaat 1080
ttctacaagg gtgtcaggac cataccatca gaaaataata tttttcaaca aatcactttg 1140
gggtcaattgc atagatacat gcaaaacaat gaagctggac tcctaacaca tactatatta 1200
aaaattaact tcaattgatt acagacatga atacaagagc taaaactaaa aattatagaa 1260
gaaaaagtaa gagtaactct tcatgacctt taatttaaca atgaattatc aataatgata 1320
ccaaaacaca agcaataaaa aataaaaaag acaaaagaca acccacaaa tgggagaaaa 1380
tatgcgcaa attacatata taataaggat ctgttatcca gattataact cttacaacct 1440
tacaccaaga cataactcat ttagaaactg ccaaaagact tgaatagaca tttcttcaaa 1500
gaagatatata tagtggccac agaagcacat gaaaagatgc tcaatatcat taatcattta 1560
gggaaatgca aaataaaatc acccatgaga taccacttta cacctactag gatggctata 1620
atcaaaaaga aaacagaaaa taataaagggt gttctcagga tgtggagaaa tttgtaactt 1680
catacactgc tggtaggaat atacaatggc acagccacca tggagaacag ttgggaagtt 1740
cttcaaaaag ttgaacagaa ttacaatatg acccagcaaa ttccactcct agatatatac 1800

ccacagaaaa ataaaacttg tgtccactaa aaccttgtac acaaatgttc acagcaatat 1860
 tattcataat agctaaaaaa aaagtagaaa caaccaatca atggataaat ggataaacia 1920
 atgtc 1925

<210> 12
 <211> 408
 <212> DNA
 <213> Homo sapien

<400> 12
 ggttcgcggc cgaggtctgg gatgaagaat tctgaagcgg gaccactacc aggtcctgcc 60
 ccctgcccac agccctccca ggccgggcag ggcaagttct ggagggccgg tgggggcata 120
 cactgaaggc tgtgtgacgt ttctatttct caaggcagta acagcaaccg tgaacctcag 180
 aggcagccaa gggaaatgtt cctcccatat ggaaagtcag aagctgccag agaggcaagt 240
 ggagcatgca agacaactga tggcatagtc tcagaactga ccatgaatac ttgctctcca 300
 ctttccattg accaaagcaa gtccaacgtt gtgggaaagg gtccctcacc cacagtggga 360
 ggtgaggggt gtggacactt gccacttgct gattgatgac caaaatat 408

<210> 13
 <211> 525
 <212> DNA
 <213> Homo sapien

<400> 13
 tgctctcggc tgcgctcttc ttgctggcgt aatgggtcac gttgatgggc ccgtgcctcc 60
 agccccagcg ggccggcaga gggccatgcc aggaccttgc aacaaacaca aggtgctcag 120
 taagtgtcga gtcctgggat gaagaattct gaagtgggac cactaccagg gcctgcccc 180
 tgcccacagc cctcccaggc cgggcagggc aagttctgga gggccgggtg gggcatacac 240
 tgaaggctgt gtgacgttct tatttctcaa ggcagtaaca gcaacctga acctcagagg 300
 cagccaaggg aaatgttcct cccatatgga aagtcagaag ctgccagaga ggcaagtgga 360
 gcatgcaaga caactgatgg catagtctca gaactgacca tgaatacttg ctctccactt 420
 tccattgacc aaagcaagtc caacgttgtg ggaaaggggc cctcaccac agtgggaggt 480
 gaggggtgtg gacacttgcc acttgctgat tgatgaccaa aatat 525

<210> 14
 <211> 504
 <212> DNA
 <213> Homo sapien

<400> 14
 ggtttccac caatttatag ccaggagtgg ctctattttc tctcgtgttg ctccagtcca 60
 ctaatttata gttctcgtgc gaagaacttt gtagctcaga aacaaaagat agagcaaaaa 120
 agagctctct cagggttagg acgtgccaca catataggac atttaaatgg ccatcttctt 180
 aataattcct ggggacatta aaactcaaat ctctggttgg aaaatttgaa aagtttgtaa 240
 accttatgtt cggcaaacgg taatagaaaa tatgttgatg atgaacatat ttggtttcca 300
 taaaaactgt gcttcccat tctaaataga tgctagtctc tctattcctc ctgggctggt 360
 aaataaaagt ggcccaaat aaaaaaaaaa aaaaaacaaa caaacaacaa aaaaaggctg 420
 ggggcggaac ccctgggcaa agcgtgcccc cggggggaaa atttggtttc ccggcacaat 480
 cccaaaatca agacaacaaa aggg 504

<210> 15
 <211> 694
 <212> DNA
 <213> Homo sapien

<400> 15
 cgtggtcgcg gcgaggcccc cccttttttt tttttttttt tttttttttt ttgggggtttt 60
 cccttttttt ttgggtttta aaatttcccc tcaaaaaata aaaatttccc gggggggggtt 120
 tcctaacacc cgggggtttt tttttatccc tcagggcctg ggggtttaaa aatttaaaaa 180
 gccttgagat tttttttaa caaaattgtg attattggcg ccagggcagg gttgcgctac 240
 aggcgctggt atccccacgc atttgtgaat gccccacacg gcgggttgtg aatgcgcctg 300
 agtgctcagg gaattattag acgacgcggt ggtgcgtcat aattttgtag aacccccggc 360
 ttcattctaa aatataccaa aaaaatttac gccggggggg ggtgcggtgc cgccttaac 420
 ccccggttat atcatctcca ggggaggcaa ggaaggcgtg cgacgggaag aatggcgctg 480
 tagacgcctg tgggaggggtg gaaacgtatg acagagcccc ccacaatttg ctcccaattg 540
 tgcccccca gccggtggca gaaaagacag gaaaccccct tcttcaaaaa aaaaaagggg 600
 agagcgttgc ggcgtactac tgggccagaa gatggcccc gggtgaaaaa tgttctcccc 660
 gccccaaccc ccataacctg aaaaaaaaaa gtcg 694

<210> 16
 <211> 988
 <212> DNA
 <213> Homo sapien

<400> 16
 accaacaac aaccacaaca ccaccacccc aaccaccggt gatagatcac tatggggcca 60

10

```

tggtgcctct agatgctgct cgagcggcgc agtgtgatgg attggtcgcg gcgaggtaca 120
ttaaaaaata tgacctcaat tttttaagtg tttaggatac aatgtaaatt acatataaat 180
caaagctctg ttttccttgc acacaccctg ggtgagagac cgccgctccc ggaggtcttt 240
cgtcctctgc agaacacacc tgggggtggg gaaaggtgtc ggctgaagca tggagcacgt 300
cctccgggct cccagtgac cttgggcact gcccccaac agagcttcag gcccctccc 360
cactatggcc ccgaggatgc ccctcccagc ctgtctgagg agtcatgcca agtcctggc 420
accaggggtt aaattccttc atttgagcac acgtccggcg gcccttcatt gtaagctctc 480
agtaaacggt tccccgaaa ttaaaatata aaaattctcc aacttcaatc catgaaatga 540
attataatta gagaaaaata aaatatgttt tagttttaat tttctataat cttaaaaaat 600
atztatgtat ctatctttta tgtctccgag aaggcacaca cagaaagtaa aaagcccagg 660
gcgggggctg cgcagcctgc ctcaggcct tcctccagca agggaggctc cccagtgcg 720
ccgcccgtt cccaggcaa ctcccagact gtgtccagtc cccaccctgg cagtctggc 780
aacaccaagc gagcttcttg aagccactaa cactcaagtc tcatactcaa catcaacaga 840
ccccggctc atgggattgt acattaaata gacatactcg aatgcatggt tgttatgctt 900
aaaaataagc taaagctggg tatctgtcaa gctgtctggt gaatgttcgc cccccaacaa 960
aaaaaaaaa attatataaa aaaaaata 988

```

```

<210> 17
<211> 221
<212> DNA
<213> Homo sapien

```

```

<400> 17
cagcggcgcc cgggcaggta ccgggcagct tgacctccat tgcttttggc ttttgtctct 60
ttctcctttt gaagctcaaa agggcataga gtggactctg atcctaggat ttttttttcc 120
ctgctttggc tgctctgtt ttggttcatg tgtcaagcag agacggggaa agccaaacga 180
cacaatgagc gttctcagaa aggaaacttc ttcggaatga a 221

```

```

<210> 18
<211> 765
<212> DNA
<213> Homo sapien

```

```

<400> 18
actagtcaca tatttattca aaagcaattt acaaagcttt ctatatcttt tcaacatatc 60
cacctgcccc tccagcatct ttttggcatt caagtaacca ctatgattta ccccgctaag 120
aaaaattatt cacatttcca tggcacaaat tgtaggaaag gaaaagacat tcttattcaa 180

```

11

gcagcggaaag ggttttggtg aaaaaacagg ttctggttct ggggagggtt ttgttatgtt 240
aggtgatcgc ctctgaagtt tgcattcatca tagagctctt aacgttaacc agggcttttg 300
cctcaggagg gaagccttat ctgtagcaga gatcagttgc aggaacacag cccaaatctc 360
aaacagtggc acccagactt gatggccgac caggcacaga tgtaagccat aaaaaagtac 420
tcatttgctt gctggctaca agaaggagca ttttatctag tgagtccatc aggaggtcag 480
gcgtaaagaa acatgtaccg ggcagcttga cctccattgc ttttggtttt tgtctctttc 540
tccttttgaa gctcaaaagg gcatagagtg gactctgac ctaggatttt tttttccctg 600
ctttggctgc ctctgttttg gttcatgtgt caagcagaga cggggaaagc caaacgacac 660
aatgagcgtt ctcagaaagg aaacttcttc ggaatgaaaa gctttggcca cattcgaaag 720
ggtagaagtc tgagagaaac tttctcatca gggagactag gtcgg 765

<210> 19
<211> 408
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (268)..(268)
<223> a, c, g or t

<400> 19
gaccaactgt gctccatctc cacgaggttg tgaagagaga aaatgggccg cctgcactac 60
agcatgagag ccatcagtta gacaaaaaga agcatggtga gacaggcaag gccctccaga 120
gaaagccagg aaggcagtga gtggctttca aaaccgatgt ggtgcattca gaggctggaa 180
ggatggacaa tattactttc ccagaaagtt tcgcaaaact ttctcttttg ttgacatgtt 240
gaaaatagca agccattgcc gtcccgntt tcccccccg gtcccgccct gtgcgtctgc 300
tggcaagcat gttaatttcc agaactcaca gaattaaagc cagagaggat ccttgtaact 360
catcttctct cctccccag cctccacag aaccataccc aaaagctt 408

<210> 20
<211> 1154
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (1014)..(1014)
<223> a, c, g or t

<400> 20
 atggtcaagc aaggcaggaa tgctcaggcc aatcggttct ggctcaaate cattccagag 60
 gggagtgtatt tacaacttag caccatcttc gtctcttccc tgagagtga gttaaatgac 120
 ccaagaataa cattagtctg caccagacct cgcgagaatg atcttcctaa ggggtggctct 180
 gggcatgggt tttcacctgc agaaactcaa gccagacat ccctccaaag ccctgtttta 240
 ctaaagcatt ttaaattctg tgggacagat gggaaaataa aacttgctgt tggaaccttg 300
 ggatttataa gaaactcctt ggtcaacatt ataaggagga cggaatcttc caagctaatt 360
 cttaacaatg caaagagtgt gatgtgtttt ggacacacac acctactcca aaatgcaagc 420
 caatgtatgc atctgtgcag aggcagacac ctacgcccat cagcacagca catgtgcaca 480
 ccgcccttcc ctgacgctc aagtgggtgag tataaggcag aggtgtctga gtcttgggca 540
 gaagcagcag gaggaggccc cagccatccc ctggcctctg gagcagaaac atggggagcc 600
 tttgaagttg ccacaactca ggtggaagcc cctcagagca gccctaagag gaagtcattc 660
 attttcaaac aaaggggaaa attcctcagg accctgggtgt ttccactaaa acaatcctct 720
 gaaaagttat tccccagagg agcaaggacc agctgtgctc catctccacg aggttgtgaa 780
 gagagaaaat gggccgcctg cactacagca tgagagccat cagttagaca aaaagaagca 840
 tgggtgagaca ggcaaggccc tccagagaaa gccagggaag cagtgaagtgg ctttcaaac 900
 cgatgtggtg cattcagagg ctggaaggat ggacaatatt actttcccag aaagtttcgc 960
 aaaactttct cttttgttga catgttgaaa atagcaagcc attgccgttc cggntttccc 1020
 ccccggttcc cggcctgtgc gtctgtggtc aagcatgtta atttccagaa ctacagaaat 1080
 taaagccaga gaggatcctt gtaactcctc ttctctccct cccagcctc ccacagaacc 1140
 atacccaaaa gctt 1154

<210> 21
 <211> 735
 <212> DNA
 <213> Homo sapien

<400> 21
 gatgattgcc atataggcga atgggcctct aaatgccatg ctcgagcggc gcagtgtgat 60
 ggatgggccg ccggtgcagg tccccccccc tttttttttt tttttttttt ttttttttgt 120
 tttaaaaaaa ttgactttgc ttttttactt tgggcggtgg ggcctgctt gaggtggtag 180
 tgtgccaggg ggatgggtgg cctgtggaaa taataccaaa agtgtgtctg aaaggaagag 240
 ggtgttgttt ttgaaggccg ggcccagggt gccctcaagt gcccgttat cttgagaaag 300

13

ggagacacgc cttgagagaa agagaattaa tgggaaacgc catacgttag gcgccaccaa 360
 ttacatgata taaaaaattc ttggaaaaat ctatgctgac catcactggg ggggtccaca 420
 gtttctcaca tcatggcggg caatggaccc cgggtccctc tctggtgtcc ttgtgggaga 480
 aggcgcagga tatgtcctgt gattcacatg agaagctggg gactgaaaat tcatgggcca 540
 ttacgcttgt tccttgggtg tgaaaatgag gtgtcatccc cgctccacaa tttccccac 600
 aaatattatg cgaaaaacaa tcggccccca ttttgtggcg acgccaacg gtgagcaacc 660
 gcaaggaaca aaaccgatac atgcaactga caaaaacaac cattcatgaa cacacaaatg 720
 aacaaaatca agagt 735

<210> 22
 <211> 218
 <212> DNA
 <213> Homo sapien

<400> 22
 catttaggcc tcgtgctcta gatgctgtcg agcggcgcag ttgtgatgat cgagcggcgg 60
 cccgggcagg tactagctct gaaaaccatt acgaagcaat gaactcatct gcaaataaaa 120
 agcacatatc ttttaatttct aatgttttat tatagatttt taaagataca tatttatattt 180
 tatattatta gcttaaagaa agtaagtcac acaagaat 218

<210> 23
 <211> 4779
 <212> DNA
 <213> Homo sapien

<400> 23
 gacactataa atgtctttcc ttatctgtgt gtactcttat ctcactgttc tattttttct 60
 cctcatttat attaaactctt tcttaccttt ttttctgaac ttctaggcct tctctttcca 120
 gaactggtgg aagacaaatg aaacggccaa gatggtaaga aacaagccgc atttctcctt 180
 ggggagactg ataatttaaa aggtttgttg tgtcagaaac attcccagct tcatcaccaa 240
 ccctttcctt ccacctctgc cactggaga ccacttacat cccgaagcgg acgcggcagc 300
 tgaagtcagg aaaccatgca tcacattagc aggagccaac tgcagacttt aaactccgtt 360
 caacatgtgg atgcggcaga gaaatgacct gtccagacaa gccggggcag ctcataaact 420
 ggttcactctg ctccctgtgc gtcccgcggg tgcgtaagct ctggagcagc cggcgctcaa 480
 ggacccggag aaaccttctg ctgggcactg cgtgtgccat ctacttgggc ttcctgggtga 540
 gccagtgagg gagggcctct ctccagcatg gacaggcggc tgagaagggg ccacatcgca 600
 gccgcgacac cgccgagcca tccttccttg agatacccct ggatgggtacc ctggccctc 660

cagagtccca	gggcaatggg	tccactctgc	agcccaatgt	ggtgtacatt	accctacgct	720
ccaagcgag	caagccggcc	aatatccgtg	gcaccgtgaa	gccaagcgc	aggaaaaagc	780
atgcagtggc	atcggtctgc	ccagggcagg	aggcttttgt	cggaccatcc	cttcagccgc	840
aggaagcggc	aagggaagct	gatgctgtag	cacctgggta	cgctcagga	gcaaacctgg	900
ttaagattgg	agagcgaccc	tggaggttgg	tgcgggggtcc	gggagtgcga	gccggggggc	960
cagacttcct	gcagcccagc	tccagggaga	gcaacattag	gatctacagc	gagagcgccc	1020
cctcctggct	gagcaaagat	gacatccgaa	gaatgcgact	cttgggcgac	agcgagtggt	1080
cagggctccg	gcctgtgtcc	tctaggagcg	gagcccgttt	gctgggtgctg	gagggggggc	1140
cacctggcgc	tgtgctccgc	tgtggcccta	gcccctgtgg	gcttctcaag	cagcccttgg	1200
acatgagtga	ggtgtttgcc	ttccacctag	acaggatcct	ggggctcaac	aggacctgtc	1260
cgtctgtgag	caggaaagca	gagttcatcc	aagatggccg	cccatgcccc	atcattcttt	1320
gggatgcac	tttatcttca	gcaagtaatg	acacccattc	ttctgttaag	ctcacctggg	1380
gaacttatca	gcagttgctg	aaacagaaat	gctggcagaa	tggccgagta	cccaagcctg	1440
aatcaggttg	tactgaaata	catcatcatg	agtgggtcaa	gatggcactc	tttgattttt	1500
tgttacagat	ttataatcgc	ttagatacaa	attgctgtgg	attcagacct	cgcaaggaag	1560
atgcctgtgt	acagaatgga	ttgaggccaa	aatgtgatga	ccaaggttct	gcggctctag	1620
cacacattat	ccagcgaaag	catgacccaa	ggcatttggt	ttttatagac	aacaagggtt	1680
tctttgacag	gagtgaagat	aacttaaaact	tcaaattggt	agaaggcatc	aaagagtttc	1740
cagcttctgc	agtttctgtt	ttgaagagcc	agcacttacg	gcagaaactt	cttcagtctc	1800
tgtttcttga	taaagtgtat	tgggaaagtc	aaggaggtag	acaaggaatt	gacaagctta	1860
tcgatgtaat	agaacacaga	gccaaaattc	ttatcaccta	tatcaatgca	cacgggggtca	1920
aagtattacc	tatgaatgaa	tgacaaaaga	atcttctggc	taggggtgta	gatatatatta	1980
tgcatttttg	gttttgtttt	taaatcaagc	acatcaacct	caagcccgtt	tagcaatgag	2040
gcagtgtaga	tgaatacgta	aaataaatga	ctttaaccaa	gtagctataa	tgggacttag	2100
cactgtatgc	atacttaaaa	aggttttgaa	aaacaaacta	cttgagaaat	atttgtttat	2160
atttttctct	aacatcatgc	tatgtgtcag	tctgaacatc	tgacaacaga	aatttcagtt	2220
attattctag	ctaagttttg	aaaacatttg	tcagtctgtt	taatagaaaa	ctgcaaacca	2280
gagatactga	ctccattaat	aaacatatt	ttgtgccgtt	ttgactgttc	tgaccaaata	2340
ctaattggga	caattcttga	cgtttttctg	ttgtgattg	ttaacataga	gcagtctcta	2400

cactaccctg aggcaactct acattggaac actgaggctt acagcctgca agagcatcag 2460
agctgaccat acattttaaac agaaatgctg gtttatttgc aaaatcacca gtatatatttc 2520
tattgtgtct ataaaaaatc agtcatttaa gtacaagaat catattttcc attccttttt 2580
agaaatttat tttgttgctc ctatggaaat cattcacatc tgacaattta tatgttaaag 2640
agttttactc tctctatttt ggtccaattt gtatctagt gctgagaaat taaataattc 2700
taaagtatga agttacctat ctgaaaatgt acttacagag tatcatttta aaatggatgt 2760
ctctttaaaa attttggttac ttttaccac aatgtaatat aatttatgta tattttatta 2820
ataatagtga attccttaaa atttgttcta tgtacttata ttttaattga tttaatgggt 2880
actgcccaga tattgagaaa tggttcaaatt attgagtgtg tttcaatata ttatctggct 2940
tatttcaaca tgagtaatat gagcaaaata agttaaacc tgcgtctgat caattttcct 3000
catgactaga actaaaacag taaatttgga caatattaag cctcaaataa tcatctccaa 3060
actccttcta acacttttta aatcagattg gaagacatgg acaaacagg ttcattgtgt 3120
gcatctttat gtcctttgcc aatatccaag atcatcacat atggtagata ttcacatgga 3180
gtttcaaatt cagaatagat taccattacc ttcctgcct tacacatcct actccttatt 3240
taaaagtctt atttgtgact tttcatttcc tgaaagtta aaaatacaat ttgagaatgt 3300
ttataataca ttctctcctg tcttttcacg gttacgtctg ttattgctga aatacaccac 3360
attttctttg ttctgggtcaa gggttaactca atatctgtgt gaaagagaac tactaacaac 3420
gttacaatag aggctagatt tgaaaaaaaa aatctataga tctaattgat acaattgtag 3480
aacaaaatgt caaaataatg ttttaagtat aagagaagat ggaccaagga gagagagatc 3540
at ttgaaaat ctaattgtag cttttctagg ctcacattca tgtactactt ttagcaccct 3600
tatgggctgt gctcgccccc tggacagttg agctttggat tatcttcctc ttcaattttc 3660
cctctattga cccgagtgtc tccctctgct tctacagatt tatagtactc cttggctctt 3720
ttgagtctcc acttttactc actgtctctg ggatttttaa gatccttttc ttctcttata 3780
aatcatcctc ttaatgaaaa ttagcctaac aaaagtttg agactggaat cctactttga 3840
gccactgact tgaaataact cttttggcaa gttgcctgac atcctgtctt accaagggtg 3900
catatttgca tttttactgc ttaaaacatt ttttttttt taccatcttt atccaaattt 3960
atcatattga tggtaggact aacaggcttt ttagaagctg gctttaactt tgagtctcaa 4020
gttacaatgc tgttgggcag cctggtcttc ccacgtgagg gtttaacttt gtttatttgc 4080
ctccagttat tccaaaatgc ttattaaatg aaaggcccag gaacatgttt attttagtca 4140
cctttgcttt ttaacaattt tgttttgtaa tcaatgagta attcatgatg aattattttt 4200

gactaatgga tagccgaagg ccaagctttt aattctaata ggtaatgttc ttcttttgtc 4260
 ttattgaaac aatgagaata ctctgtgcat ttcaaagca ctccgattat gctgtgggtt 4320
 tattcacata agcacaatat gtgttttatt tataacttca taacaaactt ataataaat 4380
 aatttacctt agcagacatg caaaagctta ttcttgtgtg acttactttc tttaaagctaa 4440
 taatataaaa ataaatatgt atcttaaaaa tctataataa aacattagaa attaaagata 4500
 tgtgcttttt attttgcaga tgagtgcatt tgcttctgta gatgtgtttt cagagctagg 4560
 tacagaggaa tgtttgctac ctttagcggg gaaaaaagaa agagagtcaa gaattttggt 4620
 ggattgtgtt tgtgtgtgca tatatttgat atcatcatta tatttgtaat ctttggactt 4680
 gtaatcatag cctgtttatt ctactgtgcc attaaatata ctttacctta tacataacga 4740
 ataaaatacc tagatgtaga tttatttaca aaaaaaaaa 4779

<210> 24
 <211> 1173
 <212> DNA
 <213> Homo sapien

<400> 24
 ttacgccaaag cttatatcgt gaaatccaag agaaaggaaa aataaaaatg tttaatgcat 60
 tttgaagcta agttgtccaa ctacattatg tctattgtcc caaatacctt ttccctgcact 120
 atgtattgat tcattgtcaa ttcattttta tacctaagcc cttacactac attacaatag 180
 ttaacctttt catcaatatt atccaatgct tggcacagaa tagaacacta agcagaaagt 240
 aaccattgtc attattacta ttgggactat cattatacat gtaaaaagat tcttcctgtg 300
 ttcaaactgt agacaagatt gaatgacaag aggttgtctt tacaccaata tttaatattt 360
 gagtctccag agtcaccata ttacaaccag ggagaattaa aacatgatat gaaattgtct 420
 tagtaatgaa tttatcacct ataaaatacc cataaaacat aactttgtta ttgacagtaa 480
 cttctgattt atccctgccc attatcta atctttttga ttgtcctaac tgatagtcaa 540
 catctagcaa tacaatgcaa gtacagtcaa tgtaaataga ttgcaaagcc gaagtgcata 600
 tctttccaaa agcatgggct ttcataaaat cagtttgggt gatttcagag aactgcttca 660
 attataggca aaggaaactc cagaagaaaa ctagttaaca aatgagttgg ataaaggaag 720
 acgatggaca cttaaatata tttggattaa agggtagaa aagagtcact gtcaaaaatt 780
 catgaagtgc ttgactattc ttttgtaaagc aggacctct ttgtgatgtt aatgttcaag 840
 tcaattgtga agagtaaggt ctgtaaagct gtcacacaat tttgtagaaa aaattaacca 900
 tttcctccaa aaaattaaca ttttattcat tttttattct aagatttagt gaagttgcta 960

ttatgctatt atgaactaca tttggataaa tataaaataa acttactctc ataatttata 1020
gctacagctt ttcattctatt cataataaaa ttttgatcac atttttagtag ggtgtaaggc 1080
cttactttta gagaaacaagt aattttacga taatgaagat ctctagtatg ttaaatgatg 1140
gtgctgctgg gcatgggtggc tcaagcctgt aac 1173

<210> 25
<211> 1301
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (520)..(520)
<223> a, c, g or t

<400> 25
ggtgcttttt tttttttttt tttttcactg gtttaaaaaa agtgacgtgg tcttttttac 60
tatggggggg gggggccggc ttgagggggg taggtgggtg cccaggggaa gtgggggggg 120
cgtgggagaa gaatgatgtg accagagaaa gggcgtggaa cggaaagggg ggggtgggtat 180
gagaagggcc aggggccaga ggggctccct cagggtcccg ctgtcgggag aagggcagca 240
gcctttggag gagagggagt ctcaagtggc aagcccatat acctatgagg ccaccaatct 300
cagaattata agaaaattct cgtggagaaa aatctcttag cgcgtgggccc cactcagcgt 360
ggtgtgggtg ctcccagttt ctcttctaca ctcagtgcgc ggtctccaga gtgagaccac 420
ccgaaggggg ctcttctctg gtgtgggtct cttgtgtgga tgaaaaaagg gcgagcctat 480
agattctgcg tgggatattc tcaagtgaag agccttttgn gggcgtatac actcgagtgt 540
gtgtcaaat agcgcgtgtg tatcccgtg gtgtgtgaaa aagctgtggg tatatctccc 600
ggcgtctctc aaaacaaaat ttctcccaca cacacaaaaa catatagccg gggagggccc 660
ccaaaggggg gaaggaaaca acaagcaagg ggcgagagaa aaaaaaggag agggagacaa 720
acacaagacc aaacccaaaag aacaacagac acaaacagc aaaaaaaaaa aaggtggaga 780
caacaacaa acaccaacca aaccacacag caaaggcaag gacacacccg aaccagacaa 840
aacaatacga aacaacaaca ccaccaactc acaacaaaaa accagcacga aacgaaacac 900
acaaacacaa acacaaaaaa aacaaacaga aacaaacaaa cacaaaacca gaacgacaaa 960
aaacaacaca acacaaggga aagacaaacg acaaaacaac acacgacaaa ccaggcccac 1020
agactaagta catatcgcag gccagcgcag acgatcaaac acaacacaca acacacacag 1080
caaaaacgag caccaccaga agacagacca ccacacacca accgaggcag aagccaacga 1140

acacagcaaa gcaacaacag tcacagcaga ctcacaaaga acagaagcat acagacacca 1200
 acagacaaca accagccacg acacaacaga gcagcacaaa acacaccaag aaggaacaca 1260
 cacatcggcc caaccacacg caacaacaac acacaccaa c 1301

<210> 26
 <211> 694
 <212> DNA
 <213> Homo sapien

<400> 26
 tggttgcacg cagccaagca cttccatcct cattgctggt tagggttgca ctgtgtgaag 60
 aatacagctg attgccccg atgcagctta tgtgcttagt gatcacccgc cagatgcagc 120
 tttgaaatth tgtgatcacc aaccagatgc agctttgtgc ttagtgatca cctgaccaga 180
 tgcagctttg tacttagtga tcaccaatth tcatctcggg ggctcttctc tgatctgtca 240
 tcctctcttc atgttaacct gctctggctt tcacggtaca gactatccct ttataaacac 300
 tgaaaaccgg aaaacaacac aaaaaaaaa aaaaacacaa acccttgggc gtcaaccggg 360
 ggtccccacg gtgttacccc ggtgtggtct aacattgtgt acccgcccca caaatthacc 420
 cccaactcat ttctcaaatc acaacacaag tactccactc accaaagact caacaatcta 480
 aaaaacacat aaatcacaat atcacagaag aggcgcctga gagaggaggc acggggagcg 540
 gaggtcggag agagagcaag acgcgcgata cggagaagga ggagggccgg gtcggggcga 600
 ggtaagga cagaggaagg aggtgaacgg agaaggagtg gagacgcagc gctgagcagg 660
 agagacgaag gggcgaggat aggaagggca ccgg 694

<210> 27
 <211> 820
 <212> DNA
 <213> Homo sapien

<400> 27
 cgaggtctcc cttttttttt tttttttttt ttttttccat ttttaaaaa agtgacttgg 60
 cttaattact atgggcgggg ggggcctgct taagggggta gggccccag ggaagggggg 120
 ggctggggaa ataataacaa aagggcgtgg aaggaagggg gggtgtgggt ttgtgagggc 180
 cggggccag ggggtccctc cagggtcctc cgctctcgtg ggaggggacc agcctttaag 240
 ggagggagtc tcctgtgggc aagccattag tcttggggcc cccaatctca gattaaagga 300
 atttttcttg agaaaatctc tagcgtgacc acttcacgtg tgggttgctc cagttctctc 360
 tcactcagtg gcggctcaga ggacaccg cgctccctca cgtgggggtct catgtgggta 420

19

```

gatggcgag caaagatctc gtgatattcc atgagaagct gtggggggga tacactcagt 480
gtggccacat aggcgtgggc cccgtggtgg tgacaatgtg gttatctccg gcctctcaca 540
attctccacc acaacattca ggccgagaca caaaaacgag cacccaacgg ggggggggta 600
caagaacaaa cagcggagca gacgagccgc acaacaaaca catcgaaaca gaaataacga 660
agacagacac caacaacagg gacaccaga gaacgaagca agcacaaaaa ccgaacaaag 720
aagaagcaag gaaggcaca ccaacatcga caaccacgaa caagacaaat gggacaaaag 780
aacacagcaa acaacagacg cccacacaca accacaccca 820

```

```

<210> 28
<211> 669
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (480)..(480)
<223> a, c, g or t

```

```

<400> 28
cgagggtcccc cccctttttt tttttttttt tttttttttt tttttgcccc ggggcagggg 60
gggggggtttt tttcccatgg gggggcccg gaaatttttt ccccttttaa aaaaaatata 120
atttttaggtg ttttggggcc cccccagg ggggtttttg caaaagggga aaggttaagac 180
aacacaagat tccgtttggg gatggtgtgt gcggcatggt tgccttcagc gtgccctccg 240
tggtccgtgg acgccccctc tacacctctt ctggggccgt gtcaacctct tgtggtggaa 300
ttttcctcac ctggtgttgt cgttggtgga ccctccatgt cgggtgtggg ggggcggctg 360
agatgccctc attggatgca gccattttcc acaatttctg gtctaaaaag ggaccgtgtg 420
agaaatgttg accccctggt gtgaaaaaga agaagagaga cagttaaatg aggaggagan 480
gggacaagac agctctcttt tccttttggg gacgcggggg ggaatagctc taagggacca 540
ctccacctgt gtgggggtgt ccttcacaaa gcgggggggg aagaccgggg cgcaatagga 600
tggtccgtgg gtggtagaat ttgtatcccg gcgctcaaaa ttccccaaca aattccaaca 660
cacaaaatg 669

```

```

<210> 29
<211> 144
<212> DNA
<213> Homo sapien

```

```

<400> 29
cgcattatga ctatatagcc caatgggtca ttagatgcat ctcgagcggc gcagtgtgat 60

```

ggatggcgag gtcaacttga tttctctctc tggttttctc tcttactgta tatttattta 120

taaaactaat tttatcctga aaat 144

<210> 30

<211> 631

<212> DNA

<213> Homo sapien

<400> 30

gcggccgccc gggcaggtcc ccccccttt tttttttttt tttttttttt ttttgtgttt 60

aaaaaaaaatgt ggaattttgt ttttttactt attggggggg ggggcctgat aaggggtgta 120

gtgtgtgcgc ccaggagaat ggttggggtc tgtgaaaata ataaaaaaaa tgtcttgaga 180

agagaaaggg gtgtggtgtg ttagaaggcc ggtggcccaa gtgggtgctc cctcagtgtc 240

tcctgctttc ctgtgagaag ggaaacacgc ctttaatgag aaatgagatg ctactgtgca 300

acgccatata cgtataggtg ccaccaattc aatatttaaa aaattctctt gagaaaaatc 360

tcatagcctt gacccaactc agctggggtg gtgggtgctc agtttctcct ctcaactcagt 420

ggcggctcag attgaacccc cgatggctc catctcgctg tctctctgtg ggtgagaggc 480

acgcatagat tcgtggatat tcacataatg aaagccttgg gggcggtaac actcgagtag 540

gcacaatagg cgtgttctcc ctggtggtaa aaatatgttt tactccgtcc tcaacaattt 600

tccacacaaa atcaggagaa acaacaacta g 631

<210> 31

<211> 618

<212> DNA

<213> Homo sapien

<400> 31

ggtcgcggcg aggtatccat ttgcctcaac ctcccaaagt gctcagatta taggagtga 60

ccactgtgcc tggccaaaaa atatttttta agcagtgact taggtatcaa atataaaatg 120

aaaagtattt tataaactgg actagaacat ttagtaaaact tccttgtttt tattttttta 180

ttttttttga gacggctctg ttctattaca tgggctggaa tacagtggga agatcacagc 240

tcagtgcagc cttgaactcc tgggctcaag caatgttctc tcctcagctc ccaagtagct 300

gggctttagt gcatgtgtca gcatgcctgg cttattttct ttttttcttt ttttcttttt 360

tttttttttt attttttttt ttttattttt tttttttatt aaaaagagca ggaggaggtc 420

atattatggt gtggcgccgg aggcgggtgt ctctccaaac ctctgggggt ccagaggtag 480

tcttctccgc cgagtgttgt gtcacaacgc gctgtcgggg gagcactcgt tggggcaaaag 540

21

agtctgtcgc ctggggtaga aatgtggttg tcgcgcgccc aaatttcgcc ccaaaaaattg 600
cgagaacaca cgagaatg 618

<210> 32
<211> 531
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (258)..(258)
<223> a, c, g or t

<400> 32
ggagactgac tcatataggc caaggtccct aatcatgccg agcggcgcca ggtgatggat 60
gcgaggcgcg gcgaggtgtt agcggctctg ctctctgat tatgccttat tctttgctta 120
tttcctttac tgagaaatgc ataatttata gttgcaaata aaaaattaat gcaggagatg 180
tgttccccac atgtactttc ttattcacat ttatgccaaa aagagattat gttatcatat 240
tgggactacg ttttatanag tcttgctctg agtttactag tccaagctat attataagaa 300
gacttttagtt ctctataac atggatcaga tttttcccaa aagatattta atgcataacg 360
caaaaaaaaaa aaaaaaaaaa aaaaagcggg ggggaaaacc ggcgcaagag cgtgcccggg 420
gggaaactgg ggtccccggg ccaaatttcc ccaaaaaatt cgcgacacaa aagtgagaaa 480
aaagagcaac acacgccagc caccaaagcc accacacaac aactactaaca c 531

<210> 33
<211> 841
<212> DNA
<213> Homo sapien

<400> 33
ggtcgcggcc gaggtcccc cccctttttt tttttttttt tttttttttt tttttttttt 60
tttttttttt tttttttttt tttttttttt tggggggggg ggggggggtt tttttttttg 120
gggggtgggg gggggggggg gttttttttt gggggggccc ggggcgcca accaccgggg 180
ggaaacaaaa aaatcatgcg cgcgccgacc cagccaccaa aagaaggga gaacaagacc 240
gaaagtgaca acaccacgc gagacgagga aagatgagga gtgatgaaag aaagaagaag 300
gggacggcga cagaagcgag acgagcggag gaggggagga cgacaaagac ccgagacacg 360
acgccacgac gaacagaccg ccgaacaaca atggagaaac acaacacaga agagaggagg 420
agcgctgata agcagatgcg atgccacaac agccgctcgc cgcccgcgga atctaagcg 480
aggaggcaag actgaaaaag aagaagagtc accacaccac ccaccactca caccgaacag 540

```

atagaaaaga cagagagaga gtcgacagag agagagagac agaaatgagg tgaggcgtcc 600
agcgcccgtg cgcggtgaga gccacaagca gagatctaca atcaatgcaa gaaccattga 660
aggcggagcg cgatacaagc aggcgagcca atacgtgact catccgcggt ggggtgtaagt 720
ctgagtgtcc tcgtcaaacc acgaacacca ccgccacaag atgatgaaaa cgaacagtat 780
cataaacaag agacaaacca agaagaggca agcaagcaca gaagagaagc gcacgcgaac 840
c 841

```

```

<210> 34
<211> 417
<212> DNA
<213> Homo sapien

```

```

<400> 34
ggtcgcggcg aggtacaagc tttttttttt tttttttttt attttttttg gggtttggag 60
ttttttttca attttttttt tccaaaatag tgacttttga aaaattttaa catcccctgt 120
tttgaatttc ccacttttca aattgaggct ttcaccacta tattgattgg gatattaata 180
ccaacgacca tagtttttgg gcatcttgac ttttctctet caaattaacc atcaacgtcc 240
tctcactgtg aatttcacga aacgacctca ttacctcttt ttaatttttt cccgtggaac 300
tttacaacaa agcaacaacg cttgtggtga tactctcagt tgctcaatac catgtttcca 360
tgttgtaaaa ttggttactc cgccactcac aattcccacc aaacaattag cgacaat 417

```

```

<210> 35
<211> 1746
<212> DNA
<213> Homo sapien

```

```

<400> 35
gcggccgccc ggcaggttt tttttttttt tttttttttt tttttggggt ttccttattt 60
tcacaaagtt ttttgtgtgt gttgggccta aaaggggaaca ggcgaggggg ggtgtattcc 120
aattttctct ctctcttttc tcaggttggg aaactctcgt gggtgcccgg gcggaattct 180
cttataagaa atatcccttt tccccccaga gattataaac caggtaagcg cattatatat 240
acccacattt tctcttataa tatagagtat agtgggctcc tatcacaata tataaaacca 300
caccttttca cacaatagc gcttttagag tgtggcatte tcatctcaca cagagtatat 360
atctctcgcg cacatatata tttttttata tatactctcg tgggtggtgt ccctattgtg 420
tgcggtataa gaacattata acgcgcacaa gcgatatata tttatatatc tctctctcgc 480
gtgtgtggca catatatgtg tggggcagat atctctctct ctcttctctc gctgtgcgca 540

```

23

```

catatatcta cggcggggga tatatatata tctctcacgc gcgcgagggg aggagacata      600
tctccgcgcg cgccttttaa tattgtgtgt gtgagacaaa gtgtggattc tctccccatt      660
atatatatat actactcccg ctgctcagac acgtgtagac acagcagtag tgtgagggga      720
gagaccccc ccgtgtgaga ggtgttctcc cccccacact atatgtctca gagatatatt      780
tccacttttt ctacttttc actatctaca aaagagagcc cccgggtgat atatcttcta      840
tcgcgcgcgc catatatctt aatatatatg atgagagagg atactgcgcg tggggtctcc      900
ccaaggtgtg tagaaccccc caagtagtgg tgggggcccc ccctaaaaaa agaggtgtcc      960
cctattatat aaaccacaaa aaagcgggcg gtgggggggg aataaacacc cccgggtggg     1020
gcacaaaaaa gcgcggtgat taaccccgcg tgggggtgtg gaaacatcat gtggggcggtg     1080
tctccccgcg gggcctacc acaaactttc cccccacaca aaatctgagt gtcaccgcgc     1140
agccacagca acacacaacc acgtgtagga aacaagacac gagacacaac agcgaacgag     1200
aagaagagag aaaagcaaac cgaagagaga tagaggaaac gcagaagaca gagacgactg     1260
atgaagagac gcaaacgaca acaacaaca aacacgaagg acaacaaca caacacacac     1320
acacaaatac cagagacgaa cgaaaaaaaa ccacgagaga caagcacgac caagacaaga     1380
aacaagagaa acgaccacag agacacacac agcgaactag acaaaagcca aacaacaagc     1440
gaaggaagaa gactaagagc acgaccgaga acgcacagaa caaacgagaa acaaaaaggt     1500
aactcaccaa caagacaccc agcagacacg agagagagaa gacaaacgac agagcaaaca     1560
acaacgaaca aaaagaccga gaagaacaaa atcggacaaa cacaacacaa gcagataaca     1620
ccaaaaacga ccatacaaaa tcccacaaca aaaaactacc acaaccaaca accaacaaca     1680
cacacaggat caagccacaa acaacacaga acacacacaa acaaagaata cgaagagaac     1740
aaacgc                                           1746

```

```

<210> 36
<211> 740
<212> DNA
<213> Homo sapien

```

```

<400> 36
cggccgccgg gcaggtagag acagtctctc tctcttgctt agctgggagt gcagtggagt      60
gatcatagct cactgaggct tgaactcctg ggctcgagca atccacctca gcctccagag      120
taggggagac tacagatgtg tgccaccata ctgagctaat ttttaaactt tcgtagagac      180
aggggtctccc tgtgttgccc aggttgccct cgaactcctg acctcaaaaa atcttctctgc      240
cctggcctcc caaagcactg ggattatagg tgtgagccat tgcgcctggt cataaattct      300

```

24

```

tgtttttagtt tgttggttta ttagacgatg gaatctctct ctcttgacca ggctagaggg 360
ctgtggtgca gatctcagcc cactgcaacc tctatctcct gagctcaagc gatcctcctt 420
agcttcccaa atagctggaa ctacaggcat gtgccatcac gtccagctaa ttttgatatct 480
ttagtagaga aggttttacc atgttgga ggggtggtct gaactcctgg ctacagtggg 540
ccacctagct cagcctacca tgagtgtgt gattacagtg cgtgagccac catgcccagc 600
ctctaaagtc tgtttgctat tcaaagtaaa tatgacatgt gtttgagtca cacaaggaaa 660
gcactaaaaa agacggtggg gggaccgggc aaagctggcc ccgggggaca tgtccccgcg 720
ccaatcccaa tgaaaagaac 740

```

```

<210> 37
<211> 687
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (499)..(499)
<223> a, c, g or t

```

```

<400> 37
gcgtggtcgc ggccgaggct acctaagcaa tcaagctggc cagctggtgc accatgggag 60
agatgatcac caaacttttc ttctctttga ggtcacacac ctagattacc tgccccagtc 120
tcccttgtag ttagatctgg ctgtgaggtt gagtttttag cagtgggata acagatggaa 180
gtttccactg gcctaacca taaattcttc cacaactctt cccactttta atcttatgcc 240
cccatgtcgt ctcttctccc agccttcttc gtctcaataa atgtcactag cacatatcca 300
gtcattcaag gaaaaacaca atggagaaaa ccctcctcaa ctaccattc cctttacctc 360
actctttccc agcatcctgc aaaatctcgc tccaaatata gctccagttt gtccacttcc 420
ctcccttttc tccagtctat aaccttggtt tactccatca ctatctctca attagactat 480
tgaaataaaa tcctacctng gaatctcaa aaaaaaaca aaacaaaaaa aaaaaaagct 540
ctcggggggtc aacccatggg gcaaacgcgt gttccccggg gggacaatgt gtttccccggc 600
ccacattccc cacattggcg caagcacacg ccgcgacgcg gccggacggc cgcgcccacc 660
cacgaacgcc caccgaggac agcgaca 687

```

```

<210> 38
<211> 148
<212> DNA
<213> Homo sapien

```

25

<400> 38
gaggatcga attatgcat gggcctctag atcatctcga gcggcgagc gtgatggata 60
gcgtggtcgc ggcgaggtac aggaactggc agccgcactg gctgccagaa acgtcagtgg 120
tgctgcccac tcggcgaaaag gttaggga 148

<210> 39
<211> 815
<212> DNA
<213> Homo sapien

<400> 39
cgccccggga aggtccctcc tccttttttt tttttttttt tttttttttt tttttttttt 60
tttttttttt tttttttttt ttttccccct ttttggtttt tttttttcca aaaaaaaagt 120
ccaaaaattc cccccccccc ccctttttaa cccccgtggg ggtgtcgccc tcccttggtg 180
gaacgaaaca aaagcgggtg gtggtcgccg ctgatgatga cgtcaaccac ctacacaaaa 240
aaaaacgggtg gtggtgatcc tgtggggcgc cccccctcgt agacatatca tcatcttata 300
taattagtta gtggtgtggc gccggagggc aggggcacac actcatcaat atctttttta 360
taatcattat tatggggggg aagaaaaaaa tcatgttacc acccccacgc ggtgtggtat 420
ccaacaacac acaaaagaag agacagttag taaaacaaca aatgagttag tgagaagaca 480
acggcaggcg tgtggtgaca gaaacaatga ctgtatgcag tcgctagtct ggagcgaacg 540
tgctgtttat gtcacctccc gcccggaata gataaaaaga tgggggtggc tacacacata 600
caggaggacg acggaggaga agagaagata ctacatcaaa caaatgggg ctgacgctat 660
tattatattc gatcggggag aagaactata tcccacaga gaagacggag ggagaagcaa 720
taacaacgac gaaacaaagc gtcacaccgc ggagagaaga aatgggcttc ccccgccaca 780
ccccccacaa ccatctccaa caaccacaac caagt 815

<210> 40
<211> 138
<212> DNA
<213> Homo sapien

<400> 40
gccagtatat gcataaggat ggtgaacagg aacatttagg agcatttgat cttatgaact 60
ggtggaccgc gagcccttag ctagacaatg agaggagaat gtacaccatg taattatata 120
tgcttgccca cgaaacaa 138

<210> 41
<211> 79
<212> DNA

<213> Homo sapien

<400> 41

tgaagataga tcatataggg cgcattgggtc actagatgca tgcgagcgg cgcaggtgag 60
 gatagcggcg ccgggcggt 79

<210> 42

<211> 887

<212> DNA

<213> Homo sapien

<400> 42

atgctggtag tggttggtt atatggtgca gcgtccagaa gtatgtgcca agctgcatta 60
 atttgaatcg gccaaactcg ctatgttaga agggatgctg ttgacgtagt atgggtgctg 120
 tcttcccgtc tctcgtctac atattgactc gcttgctgctc ggctggtctg gcctgctggc 180
 gagtagagaa tcagggtcct ctcaaaatgt gcgggttata tacgggttat ccacagaatt 240
 caggcgataa cgcaggtgaa aataaccatg ttgagacaaa aaagtgccat gctaataaag 300
 gccaggaacc cggttaagaaa gggctgaggt ttgtatgcca cgtaatatc catatggcat 360
 ccagccccc ttgagttagt catttaacaa tcaatttcgg ccgtcaaaag tcagaagggtg 420
 gggaaatcct gactaggaac ttataaagga ataccaaagg gcggtttccc ccacatggaa 480
 gcattccatc gtgcgcaatc tccatgtacc cgaccctgcc gactttaccg gattacccat 540
 gtccgtgctt atctacgctt agggaaatgg tgggcagca tatcttcatt agctcatagg 600
 ctggaagcgt aatcataagg tgacggggtg agagtacggt agcgattcaa tagcttgctg 660
 atgctgttca acagagaccc ccccggttca gcccaactgc tgccgcctta ttccggtgaa 720
 tatataagtc atgaagtcca gaccgggata aagacacgac taaatggaca gtgaaagaga 780
 gccactgggt acgcaggtta agagcaggag gaatttaggg agggaaacga gaactgtaag 840
 tggttgctaa ctatcgggat agactaaaag accgtattga gattagc 887

<210> 43

<211> 425

<212> DNA

<213> Homo sapien

<400> 43

aatgtgttgc acagtgagga cgagtttccg tgtcaatgta gctgtgacaa aggtatcaga 60
 gacgatcagg ggtatgagaa acccagctgg atcatagcaa agtattactt ggcagcaaat 120
 agtgtacctg aaatagacgt gaattgaagg agaataaga aatagaacca tgtaacatca 180
 ataaagacaa aggaaataac acacacattg accaacaaaa aaaaggcaaa gaaattagaa 240

27

gaattttacat tggaatagaa acaggggtaca tatgacatca aacacccaaa ggctaagagt 300
 tgcaaggacg agaccttata agaaagactt gaaggctact tcaactgatt cacataagat 360
 agtaacactg tgtaaaaaat aggatatcca gtcaacaaat accaaacaaa aaatacaaaa 420
 gagaa 425

<210> 44
 <211> 406
 <212> DNA
 <213> Homo sapien

<400> 44
 caggagaatc acttgaacct gggaggtgga gggtgcggtg agctgagatc acaccactgt 60
 attccagcct gggtgactga gactctaact aaaaaaaaaa aaaaaaaaaa aattgattgg 120
 ctgtgcctca ttacaaatgc ttttgatggt ggagtgtgtg tgttggaat tttttttctt 180
 ttcggggtct tcaaaatttc aagaaaagtt ggatgattgg actttggaag attacaaaaa 240
 aaaaaaaaaa aaaaaaaaaa acgcttgggg ggtacttctt ggggtgtata ggtgtgtgtt 300
 cccgtggggt ggaattgtgg ttcttcggt ctcaacaatt cccccccac aaacattagc 360
 agacgcaaac gtgggagggg gaagaggtga ggagaaagag gacata 406

<210> 45
 <211> 1267
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (358)..(358)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (478)..(478)
 <223> a, c, g or t

<400> 45
 cgtggtcgcg gccgaggttt tttttttttt tttttttttg ggggtaaatt ttttcttttt 60
 taaatgggtt attcccataa ataaaaatctc ttttccactt gaatatatta aaattataaa 120
 cactcatttt acaaatttat tcccaggtat ttacatttct cccctctccc tctcccaaaa 180
 aacgcataca ttttgghatta aatataacaa cattctcagg ctcttataaa accacctgat 240
 ttctcgtggt gtgtgcacgt ttagagaggt gtgcgaagat tggctgtcgc ctctctctca 300
 cacagagaca cactctctca gtgtggtgtg tgtgtcctcc ccccttctca ggagagangg 360

```

ggagtgtgga attgtcgccc ctctcccaca ttatacactt ttgtgtgccg tcaaagggag 420
cgcgagaata taaagcgcggt ggggggcggt ataaatcttc gtggtggtgc tcatatangc 480
gcgtgtgttt ctcgctgtgt gtgtgtgcaa caatgtgtgt gtatatctcg ccgggctcta 540
cacacaaatt ttctcacaca ccacacacac acattattct cgggcgcgcg acacaaaacg 600
caaaaaaaaa gaagaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaatgaaa aagaaaaaaa 660
aaaaaaaaaa aaaagaaaaa ataaaagaaa atcaaagaca aacagaaaaa acataaaaaa 720
agaaaaagca caaaaagaaa aaaaaaaaaa taaaagagga aaaacaaaca gaaaagacaa 780
aaaaacaaaa aagaacaaaa aaaaacagac aaagaaaaaa aaaaaaagaa aaaaaaaaac 840
aaaaaagaaa aaaaacagga acaataaaaa aaaagaaaaa cacaaaacaa cagaacaaca 900
gaagaaaaaa aaaagaagag agagagaaaa aaacaaaaag aaaaaaaaaa aaaacaaaaa 960
agaacaaaaa aaaagaaaaa aaacaaaaaa caaaaaacaa gaaaaaaaaa gaaaaaaaaa 1020
caaaaaagca aaaaacaaag aaagagaaga ggaaaaata aagagcaaaa aaacaaaaaa 1080
aaaaagaaaa atgacaaaaa acacgaaaaa acaagatata acaaaacaag aaaaagaaac 1140
aaaaagaaaa aagaagaaac acaaagaaaa acaaaaaaaaa acagagaaga aagaaaaaaa 1200
gaaaaaaaaa aacaaacaaa agaaaacaga agaaacagac gaaaaaaaaa cacaagaaga 1260
caaaaac 1267

```

```

<210> 46
<211> 239
<212> DNA
<213> Homo sapien

```

```

<400> 46
acgcagcaat acgagcatga catatggggc tcacgtaata tgtgcggtgc gtccggattc 60
tttctgcag atagatttgc ctctgtgtct tgggcgaact ccagggtgag tcgattgagt 120
agcccaaacg gtatccttac cagataaata tgcatatgat cttcgaagtt attgaccgca 180
atatcaacgt gaggactgta taatacacat tcatgaaaga tggaccttga aaacgcggg 239

```

```

<210> 47
<211> 234
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (190)..(190)
<223> a, c, g or t

```


<400> 47
 cggccgcccc ggcaggtttt tttttttttt tttttttttt ttgtggtgaa gtggttaaatt 60
 ttttttataa aaaaggttgt gttttccac agtattaaag cgggggggtat tcctagtggg 120
 ccataggcgt gttcccggtg tgtggaaatg tgtgtatccc gctcacattt cccacaaact 180
 tacgagaagn atgagagtag actaagggga aatgcgagaa gatgcatacc tagg 234

<210> 48
 <211> 964
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (364)..(364)
 <223> a, c, g or t

<400> 48
 gctttttttt tttttttttt tttttttttt tttttttttt tttaaatttt ttttttccaa 60
 aatattggcc ttttgaaaaa atttaacaat acccgtgggt gtgtgaatcc cccactattc 120
 tcaaatgtgg ggctttacac ccagataagt gtggtgggggt ataaaacaca gaacgctggg 180
 tgtttggcgc aattgtgcac ttttatctct ctcaaagtga ccatacacgt gcccaagtga 240
 attctccaga agagaacctc atatcacctc tttataattt ttctcccgcg gagaaattat 300
 aaaaagagaa aagagtcttt ggggcgtaaa cactcgctgt ggtctccaat agctgtgtgt 360
 ccncgctgtg tgtgtgacaa tgtgtgtgta tctctcgcgg ctctccacaa attttccacc 420
 acacaaacat tttcgggtga cagcaaaaag ggtgtcaaga gcgaggagag gcaaaaaaag 480
 gaaggagggc agaaccgaga gagaggcggg gagtaagcag acgacaagac agtaaaagtg 540
 aggaagacaa gaacaaagca agtggcgaag cgagcaaaag ctaggagtag gagcagcgta 600
 ctgaagatgc cattcgaagg ataagtactg cgtgtagaag aggatgcaag cacggacaaa 660
 gaacatagat aggaggctga ataactgcac gcaacgacca gccagacatt aggatgctac 720
 tggtgtagat ggagacggga ggacagagaa tgcggtgagg gcggtcgcac gaaaaccagc 780
 aacagagggg gtagcgcgca cagacagcag agaagacaga acgtaagcag tacgtgagca 840
 caaaagcagg gtaaacagcc ccaccgagcg aggagagcaa aaaagctata ctggaacaaa 900
 acaaaaaaaaa acaaaaaaac aaaaccaaga aaaaacagaa aaaaaagaaa acaccacaaa 960
 gaca 964

<210> 49

30

<211> 957
<212> DNA
<213> Homo sapien

<400> 49
cggtcgccgg gcaggtacgt gtttaatttg agtattgac aaaaagcggt tattattaat 60
tctagaatca gtcaaatga tgttctgaat agaaaataag atattcggtg gtagctgtac 120
taaggcatag actcttattc aaatgagaag taactttgct aaacaccaag ccttaatcgg 180
cattttataa taagaacatc aataccaata tttaaaataa ctgtatagcc agatatgcta 240
gcaactcgaaa attttacgaa ctaaaagtcg aacatagaag aaattgcata tccatgtctg 300
cataccccta aggatgcctt ttgggtgctg atattttttg aaaatgagag tgggtcccaga 360
aatggttcat gttgtacaag taatttgctt ccttatgttt gtttccttat ttatacacgg 420
gggtggactgg agagaagggg caaagtcaat ctgtctgtac atccgcacca gtgtggtacg 480
gtgcatcttc catgttacct ccctcttggg agatcagaca ccatatgttt tacaatacgc 540
gttgcccatg gcagtattgc ggcgaaagt gcgtttggtt tgtttcaata ggggctggtg 600
tacatggttg tctaaatata gtgtgaagtc ttcaatttct gaaggaaact aaagagacga 660
catatgtgtc ccctaagggg tctactaagt ccccatattc tctcttttgg ggctttaaca 720
gtggctagcg ggtcgagaat tcgcaagaac ttcccacgtc acgtagcttc attggtggtt 780
gtggctacct atccgatgag ttctttgtca ctttaggttt tgttcggtcc agggccgctt 840
agtagctaat ttagtcttcc taaattcctt cccctgtcc ccccaaaaac ttgtggtgtg 900
ggttttctcc ggggaatctt gggtcctcgt gtgggggaaa tgggtcccgt cgagcca 957

<210> 50
<211> 108
<212> DNA
<213> Homo sapien

<400> 50
atggtgcagg tgccggaggg tgggagaatg aagtgatgat atgagcggtc tgtctgtggc 60
ggagcttagc gtctcatggc atagctgtgc ctgtgtgaag ttgtgatc 108

<210> 51
<211> 124
<212> DNA
<213> Homo sapien

<400> 51
atggttgggg aggcgcataa gagagtgtct atactgaggt aaagaaatag ttacgaaaat 60
taacaacgga agtagtcatt ctcaatctcc taaaagggtg gagtaggatg caaagaaaag 120

aaag

124

<210> 52
<211> 598
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (469)..(469)
<223> a, c, g or t

<400> 52
gtcgcggcgg aggtcccccc tttgattatt tttttgcttt ttttgttttt tcttcatgat 60
ttgaaagacc tcgcctagat tgttttcgtg gttattgctt ggaggagca acacaaataa 120
aaagttgaga ggcccatggt gtaatactgg gggaaaatgt ggggacgagt ccaaacaaca 180
tgtgtaccgc ttttttccgg ggagaaagaa actagtagca ccttgatcc cgtcggggaa 240
cagaaatccc ctcathtagg cgcgtctgcc ctgattgccc gcaagattag tatcggttat 300
tcaagagggc acccagatta tatactacgg gaaggcgcgg tggggaggca caggtgacac 360
tggaaggcgc ctctcgctcg tggttggagc catcggtgcc accgctggcc tccacccttc 420
tccacacgca aatcttgggc agggaaaatt cctggctgtg gtctataana taacactttc 480
ttaagcatgc cacaaaaaac aaaaaaaaaa caaacaagg tctgggggaa cccctggcgc 540
aaagggtccc ggggtaacat gttgtaatcc ccgggccaca aaattcccc acaaatat 598

<210> 53
<211> 481
<212> DNA
<213> Homo sapien

<400> 53
gagcgagagg gcggagaggg gagatactat atgggcaatg gtgcttagat gctgctcgac 60
ggcgcgggtg atggatagtc gcggcgaggt acattttaaa ctagattgct agcctatgta 120
tttgacatta tcattttcag tgatgtataa ctgtcacttt ttaattttat atattatgta 180
tttatttgat attagattta ataactatat aaattttatt cattctttat ttgaatagaa 240
ataaaagttt taagagaggt tataaatcac tttattcaag tatttagtat atgataatcc 300
agttaactct gcgtagacat agatctgttt accctatcat tttcttataa taaattcttt 360
gaaattaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaacct tgggttattt cttggacaaa 420
tttttccttg tttaaaattt tttaattcgc ccaaatttcc cacaaaaatt gcaaaagggg 480
t 481

<210> 54
<211> 878
<212> DNA
<213> Homo sapien

<400> 54
tggtcgcggc cgaggtctta tttttttttt ttttattttt tttttatatt aaagccaaaa 60
gttattggtg ggggaaacct ttttgggcca ttcagggatt tcccccttgg ggaagggaac 120
ccggcgtgcg atgtggtggt aggaatcccc cgtggggtga aaacgttcgt gtcaccgtgg 180
tgcactaaaa gcagaggcac taacggggca gcggtgacag tgagaggggtg gccactcat 240
atagacgcag cccccacagg tgctcccaca gaaaatgtag ccgaggtacg tgggctccgc 300
agaagcagtg ctatttcaaa acatatgtgt ggtccccctt ggtttatgaa aatactgctt 360
acgaactatt tatagtgtag tgaataccaa aacgaaacgg tgattttgtg tgggtgtgta 420
cacaaccacg gtgccgtgtg ttgtggtctg cgtccgagtg gtcgcccggtg tgtgtgggcc 480
gaggaaggag acagactggg gcgttcgctc ctacacgccg tgtggttttg ggggtggctcg 540
ccccttctgt ggctccgac gctcaggcgt attccaggcg cgacagaaaa cccacttgtg 600
tgcgagaaat ggtagtgcca accaagttaa actgctgtgg gtgtgcgatc aacctgtgtg 660
ggggccaatg acgcgggtgg tctccggtgg tgggtaagaa atttgggttt attctcctcg 720
cttcactaa atgtctccgc aacaaacaat tttgagagtg ataccagaac aaaaaagtac 780
aactaccaa ttaactttaa ttctaagtct aacaaaagt attaccttat agaactacag 840
tcactatact tctataccta tagcgtacaa gcaaatat 878

<210> 55
<211> 278
<212> DNA
<213> Homo sapien

<400> 55
caacacactg atatcgtcta tggccatggt ctctagatgc tgctcagccc gctgtgatga 60
tataaatgta gcttgggagg agggaatgta tactggatat tgtaatgatt taattttatat 120
tcagtgaaaa gattttattta tggaattacc atttaataaa gaaatattac ctaaacaaaa 180
aaaaaaaaa aaaaaaaaaa aaggctgggg ttcttggcct gctgttcggg tgttgaattg 240
gttttccggc ccaaaattcc caaaaattcg agaacagc 278

<210> 56
<211> 123
<212> DNA

33

<213> Homo sapien

<400> 56

aaacaaaaca aaaacacgaa aagacaacac aatcttgatg ttagtcacta tatggcaatt 60
gtgcctctag atcatgcttc gagcggcgcc agttgtgatg gattggtcgc ggcgaggtac 120
aat 123

<210> 57

<211> 576

<212> DNA

<213> Homo sapien

<400> 57

tccacgacaa gctatacgag catcggtgca tcatggagca atgagagaga ctgttccacg 60
catgttgtag acgctgtttc ttgattcaca ggtagagcct tgctaataagg agatgacaga 120
gagagaggct cgcgtcggag ttccaagacg atgggtgcaag gtcgtcgttc gttgtcattt 180
gatactcctg gtttagccgc tattgcttca tcctcacatc ctatggcgta tgcgtcatg 240
gggtattagt aagtctcttt ttgatcctag tgacaagtct tcatggcctg taacactgag 300
attacttggg atcgatgggt caattcccga gagtattgag gtggacaggg gttgttaccg 360
tcgagtctctg gaagatccat cacgtagagc tcgaaaatgt ctctattaca taacgttgga 420
ctgaaccccg atataaacat cagtattggc attcccgga cgcacgggtg atacccatat 480
ggcttttgtg tccgttaa at tctattgggt tcattaagca ttgttttacc gttgtgggtga 540
acaagtgtg gttattccgg agtcaagcaa attcca 576

<210> 58

<211> 1043

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (437)..(437)

<223> a, c, g or t

<220>

<221> misc_feature

<222> (485)..(486)

<223> a, c, g or t

<220>

<221> misc_feature

<222> (497)..(498)

<223> a, c, g or t

<220>
<221> misc_feature
<222> (517)..(517)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (546)..(546)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (555)..(556)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (585)..(585)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (592)..(592)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (594)..(595)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (601)..(601)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (606)..(606)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (633)..(633)
<223> a, c, g or t

<400> 58
cgtgctcgcg cgcgagtgtga gaagtcgtat gtaaaacaga gcaggcacgt aggtcgagct 60
acgcgcagga agtacataga tacattgaca cccacgtatg gacattcccc aacgtatagc 120

```

agttcctcta gatacttcat gttgtgcgag acatgtgcc aacgaattg gtgtcgtccc 180
ccaatacact tatgagcgta taggacagta tagattggat gggacgtgag aacagagaag 240
tgaaggttat accgtatgac atggtgacat agcatggcag atattgtaga ggtccaacgt 300
actcacatct gccattagta gacggtgtag cacgtgtaag cgatggcatc aaggctagac 360
atgtaaatag tataggttcc attatgtctt gtcttgtctt gctcgggtgta taattctatt 420
gcttactccg tcgtctntcc attttgacta catgacctat ataggggatt acaccaagt 480
ttaanngtta agaaggnttg tgtaagttgc aagtggnttg ggaactgaca aactttgact 540
ccaaantatt aacannctg tgttcactt ctccattttt caaantgtgc gngnntgga 600
natacnttcc caaagaacac agggttacgg cantgaacga aaaacaaaca gaaaatctca 660
aattcagaaa acctctttca ggggggggtt ggggggtccg tggatagtgg gcagaacaag 720
aattacaagt tccacaccag ggagaattgg gagccaattt tcacaattag agggttaagt 780
ggggctgacc gaggtatatt acttggtcca gcatgtgggc acagaattgg agccaacagg 840
ctggaacaga gtttcgggga tttatataag caccttagaa gtctctggat gtcagggcaa 900
tcttgggtaa gctcaaacat tacgctaaaa ctccagggg gaaaattctt ccaggtagcc 960
taagctaggg gtaatccatt ggccataagc tggctctggg gtgaacttgg ttatccgctc 1020
ccaatcccc attaaaaa aag 1043

```

```

<210> 59
<211> 703
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (407)..(407)
<223> a, c, g or t

```

```

<220>
<221> misc_feature
<222> (457)..(457)
<223> a, c, g or t

```

```

<400> 59
gctttttttt tttttttttt tttttttttt tcctttttta aaaaattgac ttggcttttt 60
tactttgggc gggggggggc ggcttgaggg ggtaggggtgc ccgggggatg ggggggctgg 120
tggaataat gacaaaaatg tgttgaaagg aaggggggtg gtttggaagg ccgggcccgg 180
ggggcccccg ggcccgtttc gggaagggga cagccttag aggaaggaga ttcttgtgca 240

```

acgccatatg catggcgccc ccaatcatga ttaagaaatt tcctggaaaa catctacgtc 300
 tggaccatca ctgggtgggg ttgcccattgt tctcttctat cttgagcggg gccagggacc 360
 cccagggggg cctctctctg tgtctcttgt ggtgagaagg gagcgancgg caatgtactc 420
 tggtgatgac catgagaagc gtctgggggg tgtcagntcc agtggtgcac ataaccagct 480
 ggtgcccctg gtgagtgaga aaatgtgtgt ttacgccacg ctccaacaaa taccaccaa 540
 caaaatatca tggaggagac gacaacctgg ggccacacgc gtcggagtca ggcgagaaaa 600
 atcaaggcca cacaatgaaa cacaaaacga cagagaaata aaaacacaac gggaacaccc 660
 acaaaaaaaaa acaacaacaa aaacaagcaa gaaggagtgg cga 703

<210> 60
 <211> 2110
 <212> DNA
 <213> Homo sapien

<400> 60
 aaaacaaaaa aaaaaagact gatcagacct catatggcgc aattgtttcc tctcaatgca 60
 tgctcgagcg gcgccagtgt gatggatggc actggtggaa taaaacaacc tgcataTTTT 120
 actttgtttg cagatagtct tgccgcatct tgtgcaagtt tgcagcagca taggttggca 180
 tgctcagcac acacacaaca caaaacaaa aagccctatt acaagggttg ttgccctgtg 240
 gtgtatgtgc cctggtgtag catgtgcatt cgagcatatc ctgcacacat tgcacacatg 300
 ttatgtatca gacacgagac actcgatcac cgtgggtgta acaaacgcaa tgtgctaatt 360
 tgcaccatac aggcctttgt ggcgtgtcaa ctgagagtgt ggtcctataa gcttgattcc 420
 ccttgtgttt gtggacaatt tgagttacat cagagcctc cacaaattat cccacagac 480
 aagacaaata actgcatgca cagttgtggc aactaacggc tacgacagcg acataccaac 540
 cagtagagcg aaacgcgaca agaagagcga gacgaggcac gaagaggcta cacatgcaag 600
 gcacacacag accgccccac aacaccctc accaacagac agacagcgga gcacacagat 660
 ccacaggacg ccgcaacaag acacgaacac aaacgacaga cacagacaaa agcaagagaa 720
 gcagacgaga agacagcagc gcaaagcaac agagacggag agacaagaga cggaacgaca 780
 ggaggaagag aggagcagcg aggaaagaca gagaagaaag agaaggcgga aggaaggagc 840
 gacgacgaag ggggaggacg gaagagatgc gcgcggtcag atgagaggag gacagaaaga 900
 ggaggatgag ggagagagaa gacagaagaa cgaagagacg agagaccacc ccccgcgga 960
 cgagacaacg aagcgggagg agcgcgaggg ggagaacgag aagagtcggc agacggagag 1020
 ggacgagaag aaggtgacga tgcgagagcg ggtcgcgagg cggacgaggg gagagagaga 1080

gcgaagacga acgagcgcca ccgagcgaga ggaggaggag aagagagaag gaagaagacg 1140
 acgacaatgg cggccgaagc aacgaccgaa cgaagaagga gagagagcga agagacgaga 1200
 cggagagaac gagcgaggcg gaggaacgaa agaagaaaac gaggagcggg ggcagagcga 1260
 ggagaccgaa cggcgagaga agagagcgag gcacccaacg gagagagaaa caacgaacga 1320
 gagacagacg agacaagaac tcagaggcga agacgcacga cgcacagaca gagaagagag 1380
 aagacaagca gagaagcgca ccacggccag tcagcggagg cacagcccca agaaacaacg 1440
 acgggaccgc gagaacagag gagacaaatg agagcggagg ccacacgaac gacagtgaag 1500
 gaccaggacg agaccagcag caaggagaaa cgacgcatga gaacacacaa catcaaaatc 1560
 agacagacac gcagcggcac gcacgacgca cgacagcgag aagagagaca cagcagcaac 1620
 aagcatgcaa ggagcagagg acagcacgaa cgcaagcaac cagagcaaaa gcaagagagc 1680
 gcagggaaga gaaggggcca cagcagcaac agacgagcga cacagaggag aaagagacta 1740
 gaaaaagaga agacacagaa gacgcgacac ggaagacact aaagacacgc gagagaagag 1800
 aggaggagcc catgtcgaag ggagaccgaa gacgacgaaa gagaaacgaa acccgacgaa 1860
 cgcgagcgag gcgcagaacc acaacagagg atcaagaca gaggggagcaa agagcggccc 1920
 aagagcagag ccgacacgca gagaccggca gcagcaagac acacgacgcg acgacgagga 1980
 cgaggacgag agcagaagaa ccgcgacgaa cagcccacag agatcacaga gggccgacac 2040
 agcgaaggag gagagaaaag acacacgaga gaagcgacag gcaactgacga cggtaggggc 2100
 ccgcgaagac 2110

<210> 61
 <211> 3413
 <212> DNA
 <213> Homo sapien

<400> 61
 tttgggaaag aggggtcccca actcgagtgc cgcgtccccg gttatctcgt gaaatgtcgg 60
 cacatgttat gtggcccacc atgtggaaca taagccgcca ggttgggaca aataagcccc 120
 catcggccag aataagctgg attcaaatgc cggccattca ggttgaaccg agggtaggct 180
 ccaaaaatat ttcttcttct agctaattggg ctttggcaca acacacgata accaatggct 240
 ggctgctgaa atatcagccc tttggggtgg ctggaaggta agtctagctt tgggaacact 300
 agacatatat aatcgatatt tacttatatt gcattatata cagtgaagtc ccatacacct 360
 aggacatacc cgcaagcaag ctttttcatt cctgctttac cggtatgacg tcgtctaaac 420
 aaacatttca tttcagaaaa tctgcatcaa ttttcacggg ccattcacag tgcacaaact 480

gaaaagggtct tttttttttt ttttttctagc tccaccatct ctgcaacttg ccaagatgcg 540
gcaagactat ctgcaacaaa gtaaaatata cagggtttttt attccaccag tgcctcagat 600
agataggaaa aagatatgat tacgggttta atccatacat agcagcttac aatacttaag 660
atgatgaaca catggcagtc aagacaggta atttttcttc acaacagtgc atggctaaaa 720
ataaagatct aacaacgatc tgtgaaactg cactgcaacg tcaaggttcg ttcttccttg 780
accctcccc gtataatcaa atgaatatcc cctttaaaga tgaactccta ctaattatct 840
tgggcggttt cattcagctt tgcgcttcaa tccagggtatt tttgcttgga ttttagccat 900
agcatcttta acattcttat ttgcaagtcc tagataatga tctacctatg ttggtgcctt 960
gtttaatggt ctgacactac tgattttggc tctcatttca ctcttcagtg ttctgttat 1020
ttatgaacgg catcaggcac agatagatca ttatctagga cttgcaaata agaagttaa 1080
agatgctatg gctaaaatcc aagcaaaaat ccctggattg aagcgcaaag ctgaatgaaa 1140
acgccccaaa taattagtag gagttcatct ttaaagggga tattcatttg attatacggg 1200
ggagggtcag ggaagaacga accttgacgt tgcagtgcag tttcacagat cgttgttaga 1260
tctttatctt tagccatgca ctgttgtag gaaaaattac ctgtcttgac tgccatgtgt 1320
tcatcatctt aagtattgta agctgctatg tatggattta aaccgtaatc atatcttttt 1380
cctatctatc tgaggcactg gtggaataaa aaacctgtat attttacttt gttgcagata 1440
gtcttgccgc atcttgga gtttgagca gcatagggtg gcatgctcag cacacacaca 1500
acacaaaaca aaaaagccct attacaaggt tggttgccct gtggtgtatg tgccctggtg 1560
tagcatgtgc atgcgagcat atcctgcaca cattgcacac atgttatgta tcagacacga 1620
gacactcgat caccgtgggt gtaacaaaacg caatgtgcta atttgcacca tacaggcctt 1680
tgtggcggtg caactcagag tgtggtccta taagcttgat tccccttggtg tttgtggaca 1740
atgtgagtta catcacgagc ctccacaaat tatccccaca gacaagacaa ataactgcat 1800
gcacagtgtg ggcaactaac ggctacgaca gcgacatacc aaccagtaga gcgaaacgcg 1860
acaagaagag cgagacgagg cacgaagagg ctacacatgc aaggcacaca cagaccgccc 1920
cacaacaccc ctaccaaca gacagacagc ggagcacaca gatccacagg acgccgcaac 1980
aagacacgaa cacaacgac agacacagac aaaagcaaga gaagcagacg agaagacagc 2040
agcgcaaagc aacagagacg gagagacaag agacggaacg acaggaggaa gagaggagca 2100
gcgaggaaaag acagagaaga aagagaaggc ggaagggaagg agcgacgacg aagggggagg 2160
acggaagaga tgcgcgcggt cagatgagag gaggacagaa agaggaggat gagggagaga 2220

39

gaagacagaa gaacgaagag acgagagacc accccccccg ggacgagaca acgaagcggg 2280
 aggagcgcga ggcggagaac gcgaagagtc ggcagacgga gagggacgag aagaaggtga 2340
 cgatgcgaga gcgggtcgcg aggcggacga ggggagagag agagcgaaga cgaacgagcg 2400
 cgaccgagcg agaggaggag gagaagagag aaggaagaag acgacgacaa tggcggccga 2460
 agcaacgacc gaacgaagaa ggagagagag cgaagagacg agacggagag aacgagcgag 2520
 gcggaggaac gaaagaagaa aacgaggagc ggaggcagag cgaggagacc gaacggcgag 2580
 agaagagagc gaggcacca acggagagag aaacaacgaa cgagagacag acgagacaa 2640
 aactcagagg cgaagacgca cgacgcacag acagagaaga gagaagacaa gcagagaagc 2700
 gcaccacggc cagtcagcgg aggcacagcc ccaagaaaca acgacgggac cgcgagaaca 2760
 gaggagacaa atgagagcgg aggccacacg aacgacagtg aaggaccagg acgagaccag 2820
 cagcaaggag aaacgacgca tgagaacaca caacatcaaa atcagacaga cacgcagcgg 2880
 cacgcacgac gcacgacagc gagaagagag acacacgacg aacaagcatg caaggagcag 2940
 aggacagcac gaacgcaagc aaccagagca aaagcaagag agcgcaggga agagaagggg 3000
 cgacagcagc aacagacgag cgacacagag gagaagagaga ctagaaaaag agaagacaca 3060
 gaagacgcga caggaagac actaaagaca cgcgagagaa gagaggagga gcccattgtc 3120
 aaggagagacc gaagacgacg aaagagaaac gaaacccgac gaacgcgagc gaggcgcaga 3180
 accacaacag aggatcaaag acagagggag caaagagcgg cccaagagca gagccgacac 3240
 gcagagaccg gcagcagcaa gacacacgac gcgacgacga ggacgaggac gagagcagaa 3300
 gaaccgcgac gaacagccca cagagatcac agagggccga cacagcgaag gaggagagaa 3360
 aggacacacg agagaagcga caggcactga cgacggtacg ggcccgcgaa gac 3413

<210> 62
 <211> 585
 <212> DNA
 <213> Homo sapien

<400> 62
 cggccgccc ggcaggtccc cctttttttt tttttttttt tttttttttt ttggttttaa 60
 aaagtgcctt tgttttttta ctttgggggg ggggggccc atgagggggg aggggggccc 120
 aggggatggg ggggttgggg aatattcaaa aaatgtctct gaaggaaggg ggggtgtgtt 180
 gagggccggg cccgggggtg cccacggct ccgctttctg gggaagggga ccggcctttg 240
 agggaggggag ttctgggcag cccatagatt gggccaccaa tctcgatatt tagaaacttc 300
 cgtgaaaaat attcttacgc tggcccatca tgctgttggg ggtgcccagt atctcttcat 360

40

cacatggggg ccaagggacc ccgtgtctct tccgtgtggt ctctgtggtg agaaggagca 420
gctaattgttc ctggtatata ccagagaagc tgggcggggg aacgaccgtg gcgccaaacg 480
ctggttcctc ggtgtgtaga aattgtgttt accccggctc ccaatttccc ccacaacaac 540
agcgacaaac caaacgtgaa aaacagagat aaacataaag agtga 585

<210> 63
<211> 1066
<212> DNA
<213> Homo sapien

<400> 63
cgagcggccg cccgggcagg tacaaggcct tttttttttt tttttttttt ttttttttcc 60
ttgttttaaa aaagtgcctt tgcttaataa cttttggcgg tggggccccc acttgagggg 120
gttgggtccc ccgggaaggg ggggcctggg gaaataataa caaaaaggtc tggaagaaag 180
gggggtggtt ttaaagcgc aaggcccagg gtgggcccc cgggccccgc tctcgggaga 240
gggaggacac gccttgaggg aaggatgtct ttggcagacg gccatagttg gcgcccccaa 300
ttcatgttta atagaaattc cttgaggaat atcttacgct tgcccatcc cctggtggtg 360
tgcccagct tccttccatc tctgcgggtc aagggaacccc ggggtccctt ctgggtccctt 420
ttgtggaaag cgcgggacgt ctctgttttc catagaacgg cgtggcggcc gaaacaccca 480
ggggcccca taggcgtgg tcccccggg ggtgtgacag tctggtttta ccgcgctct 540
cccaaacttc cccccccca ccattgtcag cagcaaaaag tcggcccgct gggggctggc 600
gctacgatgc taaacacagg atcatcacga gaacacgctg cacaggcaac aaaagccggg 660
cgaagcaaga cacaagcca cacgaagaac gagatcagca agcaggcgac agaacaagca 720
tcgtaacgac aactgacct agcgtagatc tactgagcgt gccagatcag aggcgaccca 780
ctacaacagc ttcattactg aacacgtgag ccgatcgaca tcacagtacg ctccaaaatg 840
actaaggtca agtaacacag atacaatcga aacaagttgc tgaccgtagt tagtacacac 900
aactagatgt gaggatacta gagcaacaaa cgagtgaaac caagaacaga cacgtagaga 960
acagaagaag acgcggggga ctatacaaga cgacaccacc acgaaaagac aacaccataa 1020
agatacttag acgagcgaag cgaagcaaat acaaaagagg tacgac 1066

<210> 64
<211> 771
<212> DNA
<213> Homo sapien

<400> 64
tcgcggcgag gtcttttttt tttttttttt tttttttttt tttttaattt aaaaaccaa 60

atTTTTTTtg tgcggggaaa cctTTTTTtg ccTTTTTTtg gttttccct tttggaagg 120
gaaaccgggg gggaaggggt ggggaatccc ccggggggtg aggtgtttgc ccgttgtaa 180
aaaaaagctt tacgggggag gggcacgagg tggaggagtg gccaacaaa atatacagc 240
caaccgaggt gcccctaga aacaaatgca gccgaggggc tcgacagaca acagaatact 300
caaaaaggta gctgcgcccc cggttattat aaacaacctataaaaattca cagagttata 360
ctaatagaag cacagcgggt tccttggtgt gtggctatta taacaaccaa gagtagcgg 420
gtggttggtg tcacggtcct ccgtggtgtg cggccgagtg tggataagtg aggtgtggtc 480
tccaagaccc gggggtggtt ggggggcgct ccccttgggc gtcgcagagg tcaagcgggtg 540
tctcacagac ggggggaagt ataaaacctg gggctggaaa gagggcccgga aggaagagcg 600
tggggggggtt aacctagggg gggcagaata gcgagtgggtc tccggtgggg agtgaactgt 660
gtggtctctc tcggcgctcc acaacctccc cccaacattc ccggctacga caaaaaagag 720
taaaaaaaaa agaaaaggac acaagaaaa aaaaaacaa tagacgaagg a 771

<210> 65
<211> 389
<212> DNA
<213> Homo sapien

<400> 65
atggggcggtg cgactcgtcc cacaaggaag gatgtgtag aaacctgcca ctagacagag 60
ggaggagaaa gtgaagaagg cggtcagcag acaggagaag agcaagcgggt tccctcagag 120
gagtgaacgg tgctagtacc atcagagtgg accatagcac tcaagccctg acaccatgtg 180
gaaagcatta acacagatgg acaagacatc acaaacatg aaccctacgt gagttgcccc 240
aattcttttt gtaatataaa cttgggtgca atcccaacca acactcatca cctggaaacc 300
tagtatataa gcccagaaca agggccccc aaaggggcc aaccactat catacctctt 360
gtaaattaaa agaccttgag atcacaatg 389

<210> 66
<211> 843
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (415)..(415)
<223> a, c, g or t

<400> 66

42

gggcaggtac ccaggacaca aacactgctg aaggccgcag ggtcctctgc ctaggaaaac 60
cagagacctt tggtcacttg tttatctact gaccttccct ccactattgt cctatgaccc 120
tgccaaatcc ccctctgcga gaaacaccca agaattgatca ataaaaaaaa aaaaaaaaaa 180
gtttttcaac ctttgtgtta agagcccact caagagttgt tggtgttttag ctttctatta 240
tatttggtaa attttttcag tttttttttt tggcttttac tcggttgat tctcctcat 300
tcccattttg ggctcattag acagtgttag tttctcatag gaaattttcc ttttaataa 360
aatttggtgac taagcactcc ctttttggtt ccctaagagt ggtggctctt ccagngggaa 420
acagtgtgtg gtaagtgagc actattggac gaagggggtg gtgtatctcc gtgagtgtg 480
cgtcgagagg aggtgtctcc ccaataacct cgtgctcggc gaactacctg gcttttaata 540
cgtgccttaa agatagtctg ggtcctcttt atttaccatt ctcttctctt gggttttcat 600
ttttatttca ccaaaagggtg gggcggtttt caaaattttt gtggcttcta tctcgaaaga 660
aaaaaaaaac aaaaaaacct ctgtggggcg tcgaaccctg ggggcccaa cgcggttccc 720
tggttggtga aatttggttc ccgcgcgcc ccaatcccc cacttcctc ccacacacaa 780
acaaggggca gaacgacaag aaaaagaaga acaacaagaa aagaaaacaa aagaaagtaa 840
gtg 843

<210> 67
<211> 2336
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (429) .. (429)
<223> a, c, g or t

<400> 67
cacttacttt cttttgtttt cttttcttgt tgttcttctt tttcttgctg ttctgcctt 60
tgtttggtgt tgggagggaa gtggggggat tggggggcgc ggggaaccaa atttcaacaa 120
ccagggaacc gcgtttgggc cccacgggtt cgaccgcca cagcggtttt tttgtttttt 180
ttttctttcg agatagaagc cacaaaaatt ttgaaaaccg cccaccttt tggtgaaata 240
aaaatgaaaa cccaagagaa gagaatggta aataaaggag acccgaacta tctttaaggc 300
acgtattaaa agccaggtag ttcgccgagc acgaggttat tggggagaca cctcctctcg 360
acgcagcact cacggagata caccaccccc ttcgtccaat agtgctcact tacacagcac 420
tgtttccnc tggaagagcc accactctta gggagccaaa aggggagtg ttagtcacaa 480

atatttattaa aaaggaaaat ttcctatgag aaactaacac tgtctaataga gcccaaaatg 540
ggaatgagga ggaatacaac cgagtaaaag ccaaaaaaaaa aaactgaaaa aatttaccaa 600
atataataga aagctaaaca acaacaactc ttgagtgggc tcttaacaca aaggttgaaa 660
aactttttat ttattttttt tattgatcat tcttgggtgt ttctcgcaga gggggatttg 720
gcaggggtcat aggacaatag tggagggaag gtcagcagat aaacaagtga acaaaggctc 780
ctggttttcc taggcagagg accctgcggc cttccgcagc gtttgtgtcc ctgggtactt 840
gagattgggg agtgggtgatg actcttaatg agcatgctgc cttcaagcat ctgtttaaca 900
aagcacatct tgcaccgcc ttaatccatt taactctgag tggacacagc acatgtttca 960
gagagcacag gggtggggat aaggtcacag atcaacagga tccaaggca gaagaatttt 1020
tcttagtaca gaacaaaatg aaaagtctcc catgtctact tctatccaca gagaccggc 1080
aaccatccga tttctcaatt ttttccccac ttttccccct tttctattcc acaaaaccgc 1140
cattgtcatc atggcccggt ctcaatgagc tgttgggcac acctcccaga cggggtggtg 1200
gccgggcaga ggggctctc acttcccagt aggggcggcc gggcagaggc gccctcacc 1260
tcctggacgg ggcggtggc cgggcggggg gctgaccccc ctacctccct ccagacagg 1320
gcggctggcc aggcagagg gctcctcacc tcccagacgg ggcggcgggg cagaggcgct 1380
cccatctcag acgatgggcg gccgggcaga gacgctctc acttctaga tgggatggcg 1440
gccgggcaga gacactctc actttccaga ctgggcagcc aggcagagg gctcctcaca 1500
tcccagacga tgggcggcca ggcagagacg ctctcactt ccagacggg gtagcgccg 1560
ggcagaggct gcaatctcg cactttggga ttacaggtgt gagccaccgc gtccagcctt 1620
tctttttact ggttctaatt attattattt tttattttac tagtccttgc ctgcatacat 1680
ttctccagg gtacagagct tatgtggttc ttgaccaa tactgttcta gtcattgcac 1740
gtattagaga ccaaggcttt cctcgtcaaa tcaattctgc atggttttcc catcttcttg 1800
gttttctttt tttttttttt ttttttaatt ttttattgat cattcttggg tgtttctgc 1860
agagggggat ttggcagggt cataggacaa tagtggaggg aaggtcagca gataaacaag 1920
tgaacaaagg tctctggttt tcctaggcag acgaccctgc ggcttccgc agtgtttgtg 1980
tcctgggta cttgagatta gggagtggg atgactctta aggagcatgc tgccttcaag 2040
catctgttta acaaagcaca tcttgaccg cccttaatcc atttaaccct gagtggacac 2100
agcacatatt tcagagagca cggggttggg ggtaaggctc tagattaaca gcatcccaag 2160
gcagaagaat ttttcttagt acagaacaaa atgaagtctc ccatgtctac ttctttctac 2220
acagacacgg caacaatctg atttctctat cttttcccca cttttcccc ttttctattc 2280

cacaaaaccg ccatcgatcat catggcccgt tctcaatgag ctgttgggta cacctc 2336

<210> 68
 <211> 836
 <212> DNA
 <213> Homo sapien

<400> 68
 gagcggccgc cgggcaggtc cccccccct tttttttttt tttttttttt ttttgttttt 60
 tttttttttt tttttttttt ttataaaaaa aaaccggggg aggggggggg ggggggggga 120
 aaaggagaaa aggggggggg ggtggtgtaa aaaaaggggg tgtggtgcgg gggaaagggg 180
 gtggaaacgt ggcgtgcacg aggggggggg ggggtggagg acccccagc tgtggggggg 240
 tgggtgcatc aatacatgat acggaggggg gtaatacggg ggcgccaac acagcatgtg 300
 gtggaccctc atggtagaca gaggaggaga ggagttcaat cattcgagca gacgaacgaa 360
 aaacagccgg tgttcacacc accaagaaag tgtgtctccc ccacgagggg gatacaacag 420
 cggggggagg cagcgggcga tgtccgcagc gggggctgct gggaaagaaa agtcctacca 480
 caaaaccagt cccttgtggg ggggaagaag agagtgtagc agccgctcct ctgagagaga 540
 gagaaagtat atcatagcag cgagcacgag cggaggagag agagcgctc gcacaaagaa 600
 gtgaggtgag cggctgccgc agcgcacaca aaataaataa gaggagggtg tttaaacacgc 660
 cggggggggc agaaaatata acaatagtag cggcgccgcg cgagaaacaa aggtggggga 720
 aacaacacgg tgggacacca acagaggcta tccaccgagg gggtgaaaaa aagtgttttt 780
 cggcgggcca caaccatccc caccacaaac tggccgcaac aacaacaaca aggtct 836

<210> 69
 <211> 411
 <212> DNA
 <213> Homo sapien

<400> 69
 cgtggtcgcg gcgaggtttt tttttttttt tttttttttt ttttttaagt ccatgggaaa 60
 ggggtttttt tccccaaaa aattgcaagg ggaataaacc ccatttttcc aaaggcgaag 120
 gtacggcatt ttttaaggtat tccgggcctc ccttgggaga agcaaagcga ttttaaaaaa 180
 gtttggcagc gcgtgaaact cgtggttggg aacattccca ggttaaaaat atttgagcaa 240
 aagagctttc ttttaaaaaa ccacacacac ttacacctt tactagaaaa ccaagagtgg 300
 ggggttaact ctgtgcacat agcgtgttcc gcgggggtga aagtcgttta ctccgcctca 360
 caattccccc acaacatctg aggaggacag gggttgtgcg acgcgagcaa g 411

<210> 70
 <211> 1343
 <212> DNA
 <213> Homo sapien

<400> 70
 cggccgcccc ggaggtacc accattgtaa ggaaacactt tcagaaattc agctggttcc 60
 tccaaaaaaa aaaaaaaga ctgggcggta atcatgggtc gatagcgtgg ttctccgtgg 120
 ggtgaaatgg gttagtccgc tcgacaattt ccaccacaac atacgagcca aggacaaaga 180
 agaagaacac aaagcaaac acaccagagg ggggaaacaa agaaaaaga acagaccaca 240
 gaacagcagt aaacagagca caaacatata acacaacacg cagaaaagac gagaaacaac 300
 aagacagaag cgcacgcaaa acaaacgaca aaacaacaaa aaaatagcaa aaaacagaaa 360
 gtgacggccc gtcgaagcaa gagaaaggag aaaagggaaa gagaggcgga agtgagcgag 420
 aaggagaaga ggagaaaaga gagcagataa ggagaacaga ataaacgaag aagaaaaaaa 480
 aaacagacgc agaaagagga gagggcaagg agaaagaaga agaagagaga gaggatagcg 540
 cgacgagcgg aggaagtaaa gacagacggg gagactgaag aggaggaacg gagagacatc 600
 ggcacataga cagaggagga ccgccgggat acaagaaaaa aggaacaaac ggaagattga 660
 gaaaatatga cgagcgacga agcaacgacc gaaaccagac cagcgcgaga ggcagagaaa 720
 ggagcagaga acaaaaaagc gacagagaaa ggcaagacga aaaagacaag cacgagctac 780
 aggaggagcc aaagaatgag aaaagagaga agaagaaga aacacgaagc aacaagcgc 840
 agaacaggag aagagagaga aaacagaggg agacgaagag agcagaggag aagaagaacg 900
 aaagtaggga gccaaagaaga aacgaaacga gaagtacaaa cagaacaggg aagaaagaga 960
 ccaaaaggac aaaagaagga aacacagaga agaaaaaga gaagaaaaa gaaaagccag 1020
 agaagaaga caggcaaacg aaagcaagaa gaaaaaacga cacaacgaga gagaagagaa 1080
 aaagacaaga gaagcaggag agaattgaaa tacgcagaag aggaggaaac agataacgaa 1140
 gagagaagac gaaagaagag aaaaagacag cagaagaaaa gagagaagaa gagaagaagc 1200
 aagaaaagca gaagcaagaa cgaagcagac aaagagagag cggaaacgac aagaagagaa 1260
 gaaagagaaa gagagacaga agaggagaag acgaggaacc ggagctgagc agacagaaga 1320
 gagacgaaca gaacaaacag aca 1343

<210> 71
 <211> 3259
 <212> DNA
 <213> Homo sapien

<400> 71

atgagcgcgg acgcagcggc cggggcgccc ctgccccggc tctgctgcct ggagaagggt	60
ccgaacggct acggcttcca cctgcacggg gagaaggga agttgggcca gtacatccgg	120
ctggtggagc ccggctcgcc ggccgagaag gcggggctgc tggcggggga ccggctggtg	180
gaggtgaacg gcgaaaacgt ggagaaggag acccaccagc aggtggtgag ccgcacccgc	240
gccgcactca acgccgtgcg cctgctggtg gtcgaccccg agacggacga gcagctgcag	300
aagctcggcg tccaggtcgg agaggagctg ctgcgcgccc aggaagcgcc ggggcaggcc	360
gagccgcccg ccgccgccga ggtgcagggg gctggcaacg aaaatgagcc tcgcgaggcc	420
gacaagagcc acccggagca gctctccctg gtggcagtgt ctgatgggag tgtccgtggg	480
gctacgagga gcctcctgga cagagaaagg gcacagttcg gcattaagag gcagaaccca	540
gccctgcccc agcttgggcg tgaggggtcca agagcaatgg tggctgagct cggccagcgc	600
gagcttcggc ctcggtcttg taccatgaag aaggggccca gtggctatgg cttcaacctg	660
cacagcgaca agtccaagcc aggccagttc atccggtcag tggaccaga ctccccggct	720
gaggcttcag ggctccgggc ccaggatcgc attgtggagg tgaacggggt ctgcatggag	780
gggaagcagc atggggacgt ggtgtccgcc atcagggtg gcggggacga gaccaagctg	840
ctggtggtgg acagggaaac tgacgagttc ttcaagaaat gcagagtgat cccatctcag	900
gagcacctga atggtcccct gcctgtgccc ttaccaatg gggagataca caaagacccc	960
ctcaccccat cctctgacaa cccacaaccc tctctctct gccaggagaa cagtcgtgaa	1020
gccctggcag aggcagcctt ggagagcccc aggccagccc tggtagatc cgcctccagt	1080
gacaccagcg aggagctgaa ttccaagac agcccccaa aacaggactc cacagcgcgc	1140
tcgtctacct cctctccga cccatccta gacttcaaca tctccctggc catggccaaa	1200
gagagggccc accagaaacg cagcagcaaa cgggccccgc agatggactg gagcaagaaa	1260
aacgaactct tcagcaacct ctgagcgcgc tgctgccacc cagtactgg cagggccgag	1320
ccagcattcc accccacctt ttttcttct ccccaatact cccctgaatc aatgtacaaa	1380
tcagcaccca catccccctt cttgacaaat gatttttcta gagaactatg ttcttccctg	1440
actttagggg aggtgaatgt gttcccgtcc tcccgcagtc agaaaggaga ctctgcctcc	1500
ctcctcctca ctgagtgcct catcctaccg ggggtgtccct ttgccaccct gcctgggaca	1560
tcgctggaac ctgcaccatg ccaggatcat gggaccaggc gagagggcac cctcccttcc	1620
tcccccatgt gataaatggg tccagggtg atcaaagaac tctgactgca gaactgccgc	1680
tctcagtggg cagggcacct gttatcctga gacctgtggc agacacgtct tgttttcatt	1740

47

tgatttttgt taagagtgca gtattgcaga gtctagagga atttttgttt ccttgattaa	1800
catgattttt ctggttggtta catccagggc atggcagtggt cctcagcctt aaacttttgt	1860
tcctactccc accctcagcg aactgggcag cacggggagg gtttggttac ccctgcccac	1920
ccctgagcca ggtaccacca ttgtaaggaa acactttcag aaattcagct ggttcctcca	1980
aaaaaaaaaaa aaaagactgg gcggtaatca tgggtcgata gcgtggttct ccgtggggtg	2040
aatgggtta gtccgctcga caatttcac cacaacatac gagccaagga caaagaagaa	2100
gaacacaaaag caaacacac cagagggggg aaacaaagaa aaagaaacag accacagaac	2160
agcagtaaac agagcacaaa catacaacac aacacgcaga aaagacgaga aacaacaaga	2220
cagaagcgca cgcaaaacaa acgacaaaac aacaaaaaaa tagcaaaaaa cagaaagtga	2280
cgccccgtcg aagcaagaga aaggagaaaa gggaaagaga ggcggaagtg agcgagaagg	2340
agaagaggag aaaagagagc agataaggag aacagaataa acgaagaaga aaaaaaaaaac	2400
agacgcagaa agaggagagg gcaaggagaa agaagaagaa gagagagagg atagcgcgac	2460
gagcggagga agtaaagaca gacggggaga ctgaagagga ggaacggaga gacatcggca	2520
catagacaga ggaggaccgc cgggatacaa gaaaaaagga acaaacggaa gattgagaaa	2580
atatgacgag cgacgaagca acgaccgaaa ccagaccagc gcgagaggca gagaaaggag	2640
cagagaaaca aaaagcgaca gagaaaggca agacgaaaaa gacaagcacg agctacagga	2700
ggagccaaag aatgagaaaa gagagaagaa gaagaaaaca cgaagcaaca agacgcagaa	2760
caggagaaga gagagaaaac agaggagagc gaagagagca gaggagaaga agaacgaaag	2820
tagggagcca agaagaaacg aaacgagaag taaaacaga acagggaaga aagagaccaa	2880
aaggacaaaa gaaggaaaca cagagaagaa aaaagagaag aaaaaagaaa agccagagaa	2940
gaagaacagg caaacgaaag caagaagaaa aaacgacaca acgagagaga agagaaaaag	3000
acaagagaag caggagagaa tggaaatacg cagaagagga ggaaacagat aacgaagaga	3060
gaagacgaaa gaagagaaaa agacagcaga agaaaagaga gaagaagaga agaagcaaga	3120
aaagcagaag caagaacgaa gcagacaaag agagagcggg aacgacaaga agagaagaaa	3180
gagaaagaga gacagaagag gagaagacga ggaaccggag ctgagcagac agaagagaga	3240
cgaacagaac aaacagaca	3259

<210> 72
 <211> 762
 <212> DNA
 <213> Homo sapien

<400> 72

48

```

cgagcggccg cccgggcagg tacgcctgta gtcccagcta ctcaggaggc tgaggcagga      60
gaattgcttg aaccaggag gaagagggtg cagttagcca agatcatgcc acatcactcc      120
aacctgggca acagaacaag aaccatctc aaacaaacaa acaaacaaaa aaaaaaaaaa      180
tctggtctcc ttaggatata gttaccgtgc cccacgtgca gactagaaga aattaactgg      240
tgttttggaa cttttttacg tgcaaaactt gaaaatgtgc tagaaacca agcattgaag      300
aattaaatta ctgtgggtgg gaaacacacg ggcattgtgc attattgcat tattaccttg      360
ggtaggttat agtaagggtt agaaaggcat agcttgggtg gatattctga accaccattg      420
aattcttttg gggccagggt tagggaaggc acagccagat tccttatggg aattgaatta      480
cctcaaattc ggggtgggtc tgagatttct agagatttaa cccactgtgg tgccattttt      540
taacaaaaaa aaaaaaaaaa aaaaaaagg gcggggggga aaccggggc caacgcgggg      600
accgcgctgg ggtggggaaa ggtgggttac cgccggcgcc acaaattccc ccaaaatttc      660
atcgcagcac aaaaaaacg aacacaccga acagacacag agacacaacg accacacaga      720
ggacagaaca caaaaggaa acaaacacac acaaagagga gc                          762

```

```

<210> 73
<211> 989
<212> DNA
<213> Homo sapien

```

```

<400> 73
gctcctcttt gtgtgtgttt gtgttccttt tgtgttctgt cctctgtgtg gtcgttgtgt      60
ctctgtgtct gtcggtgtg ttcgtttttt gtgtgctgcg atgaaatttt gggggaattt      120
gtggcgccgg cggtacccca cttttcccca cccacgcgg gtcgcccggt tggccccggg      180
ttcccccccc gccctttttt tttttttttt ttttttttgt taaaaaatgg caccacagtg      240
ggttaaatct ctagaaatct cacgaccac cgaatttga ggtaattcaa tccccataag      300
gaatctggct gtgccttccc taacctggc cccaaaagaa ttcaatggtg gttcagaata      360
tccaccaag ctatgccttt ctaaacctta ctataacct cccaaggtaa taatgcaata      420
atgcacaatg cccgtgtgtt tcccaccac agtaatttaa ttcttcaatg cttgggtttc      480
tagcacattt tcaaagtttg cacgtaaaaa gggtccaaaa caccagttaa tttcttctag      540
tctgcacgtg gggcacggta acatatacta aaggagacca gagttttttt ttttttgttt      600
gtttgtttgt ttgagatggg ttcttgttct gttgccagg ttggagtgat gtggcatgat      660
cttggtcac tgcaacctct tcctcctggg ttcaagcaat tctcctgcct cagcctcctg      720
agtagctggg actacaggtg tgagccacag cgctggccc gagagtgtgc gatatgctcg      780

```

49

caggaagtat ttctgtgtta aaaagttgag aagatggaaa ctgaatcctc ttgtattca 840
 gaaggctggt tcggaaggtc actgttggca ggcttctccc tacagggatt cagcagtgag 900
 ggagcagagt atttggggga caactgcttc ttctggaggg gcgagaatga gatggagtcc 960
 accagcagct ctttatgtca gacttttag 989

<210> 74
 <211> 1725
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (83)..(83)
 <223> a, c, g or t

<220>
 <221> misc_feature
 <222> (218)..(218)
 <223> a, c, g or t

<400> 74
 tggtcgcggc cgaggttttt ttttttttta tttttttggc agtttttaaaa aggaggattt 60
 atttggacaa gttccacttt agncgcaata tattccccta aaaggaaatc tcacaattac 120
 aagtgaaga tttaaacttc agggccctca gaattttctca ttacaaacac ccaagaccaa 180
 aatctcctag agatatctcg gtattgtgcg ttcctcanaa tttttctccc attataacct 240
 ttaaacaaac aaaagccgtg tgggcgttta aatccatgtg gtccatatag ccgtgtgttc 300
 ccgtgtgtgt gtgaacattg tgtttactcc gccctccaca attctccacc acacacacca 360
 ttacgcgagc acagggggaa gtgaaaggtg tagaaaacgt agtcggggga aagatagaaa 420
 cgacagagca aacacgcaga gctactagaa gagaaaaatc agagagaaag ataaccatcg 480
 cgtcaacgac ctgcaggaga gcagagacat ccagagcgca cgcgcggacg aaatagacga 540
 gatatgccat acgggacaca gcgcgacacg agggatatga tgagaccagc ccaactgaaa 600
 gcatatgata agagagcgag actgatacgg acgaaaaagg acgcaaacca cctcgcggca 660
 cccctgaact aaagacaaaa agaaaggaag aaaaccaaaa catataaaga aagacgagac 720
 agacgaggaa caaaaaaag aaatgaagaa gagagcaaga acgagcagac gataaatgag 780
 agaaaacaaa gatagacaga agagagatgg agagagagag agcgaagaag cactatcaac 840
 agatagacac tcacgataca gcgatacaga ataagaacta ggagacgaac gagagaggac 900
 agagaaacga gaggaacaag ctagaaccac aagagagata aaaagaaaga agagctaagt 960

50

gaacacgcag cgcgggagat ggtagagaca aagaacgcga tgacagaggg aggcgagggg 1020
actacggata gagtgagcgc agcggctatg gataaggaag atagcggata cattgagggg 1080
gggcgcgcgt aggtatcaga ccgcgagagt cattgatcga ataggaggag tacgaaggac 1140
agagagagtg atagagtgat aaaaaatcaa agataagata tacgatagat cgaggatcac 1200
gaaacgaaag agaattgcgag agaagagggg ggtagataga gagactagag agagacagag 1260
ctaagaaaca agaaaagaac aagatagata gcagtagaga gcggagcagc gataagaaag 1320
aacaaggacg aaaaagagaa gagcagagag caaagaagta acgtacgcaa acgagagaca 1380
aacctgaaca gacgcagaag acggacagcg aacaaatcga acgagtggag gaggaactaa 1440
cacgaaccca tagaagagag caaagcaagc acgcgaaggg agaaagcgac gagcaggaga 1500
gacggacacg aggagcgaga tagagggtgta atattcgca agtagcgaga agactgaaag 1560
tgaacgggcc gaggaatgaa gttaaagagt cgactagaac gacagaggac gcgaaagagt 1620
aagacatagt cggctcaagg cagtagtgat atagagcgta gagcagagga gagagtataa 1680
tgtggtccag gagcgatgag agcggacgct gagtgcgtag tatat 1725

<210> 75
<211> 1075
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (346)..(346)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (390)..(390)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (522)..(523)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (534)..(534)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (538)..(538)

<223> a, c, g or t

<220>

<221> misc_feature

<222> (598)..(598)

<223> a, c, g or t

<400> 75

```

ctgtgcgccgc ggccgaggtc acatcgtatt ctgtgccgag cttcgcacac ctgacgtacc      60
gagcatcatg atcgtctccc agcgaccctc acagattctc gggcctgcaa ccctgtctat      120
tgacgtttga atggaatggc ctgtgtcatg cactcacaga tcgctattac taccctcgtg      180
caatgaaggc caatgtgtgc gaccagatcc ttcctatgct aactcgtaag tagaatcggg      240
gtagtaactc gcgaatcacc cttagtatat ggagagacct ctattcatcc acacatgcca      300
ctactcgact tggaagaatg gcctttgttg gggatatccc gcgcgnagtt gccaaagata      360
ggtcctattg gggccagttg agagtacgan ttcgagtatc gattcacgac ctagtcttat      420
tcccgtagg tagatgggaa acaatataga tttcaatccc cagccacgag caacaatttc      480
gcaaacgagc cacaccgata tgggaagcct aaaaccctgt gnnnttccca tgnagtncc      540
caacgtttta tgttttttcc ttatttaatg tgtgaagaag ataaaaatta gtccgtgnta      600
cttcttaaaa agagagaaa agacaaagag agaaaaaaga aaaaaaaaaa aggcgtgtgt      660
gcgcgggtgt acaccccgag tgcgcgtccc aatacgctg tggtgtctcc ggtgtgtgtg      720
tgtgcacaca tgtgtgtgta tatctcgcgc gcgctccaca aattctccca caccaaaaca      780
atthtcgttg ttagaacaaa aaattgtaaa aaaaaacaaa aaacaaaaag cacagaacaa      840
acaaaaaaaa caaaagaaa aacaaaacac aaaaaaaaaa agaaaaaaaa aagaagaaaa      900
gaacaaacgg aggaaggaaa gagaagagaa aaagaggaag aagaggaata aaacgaggag      960
cagagaaaag gaccaacgca aatgagacgc aaagcacaaa caataagaga caagaggaaa     1020
aaaaaaaaag aaagacgcaa agaaagcaaa agcgaacgag gagacagaaa ggcac          1075

```

<210> 76

<211> 491

<212> DNA

<213> Homo sapien

<400> 76

```

ggctgtggtg caggtgtgtg tgcctcaact attgccaatg tgttccacct agctggactt      60
tccttccttc tctaattgat gtgcagtatg actcccatga aaatgatgaa ccttgtcatg      120
aagttctcat cgccaacgaa gaacgactgc ataggaagaa tatgaagaaa tagctgctaa      180

```

52

actgactaag atcgacttca tgtagttgaa gaaatgctct gttcacccgat ggatgccttg 240
 ctgtctctat taattgatct aaacctgttg agcagtcaga gtcttgccact ggatttagtt 300
 tagcgtgccc ataggatgca tcgcactcttg gcttactctt ggtcttagct gtttcgctgt 360
 gtgaaatcgt tatccgctca cgattccatc acaacatgcg gatgcagcac gatatactgc 420
 actagataaa tggaccaacc aactaaattc tctcaaccag gctgtagtca gtaaaactggc 480
 ttaacagaga a 491

<210> 77
 <211> 1440
 <212> DNA
 <213> Homo sapien

<400> 77
 aagaagatcg actactatag gagccatggt tatctagatg catgctcgag cggcgcattg 60
 tgatggatag cggcgcccgg gcaggtcaat ctttaaattc agtattcagc ttccaaagat 120
 ggggtgccc taatagactt aaacatataa tgatggctac agaacaata agtatacgac 180
 aaatgtaaaa acaggaaatg taagctccac tctcaatctc ataccaaggg tgagagttac 240
 gagatgctaa agcaaaataa atgtaggttc ttattatatac tatttctgt atatcatgca 300
 gtctgcttct tttagatg ccttacggag ttaccaatt taagcttacg aggattgtaa 360
 gtgcaattgg ctgggaactg acaacatgtg atccaagcta ctacaccct gtgctcactc 420
 tgtcacttct caaattctgc gcgctggaac acattcaca gaacaacagg gctagagcac 480
 tgcaaggaaa ccacacacca ccaaactcaa aactcagaaa caccacatc tccagagagg 540
 cacagagagg atacaaagaa tactgcgcaa gacaaagaaa tccacagact ccacacccca 600
 gggcacaacc tggaacgcaa aattcaaaaa actaacgcgg accaaacgcg gaagccccga 660
 cccaaaagag cacaatataa agaggcccca cgcccacgt gcgcgcacca cagcacacc 720
 cgcacggcca acacacagg ccctgaaaca cacagcacta cacacggggc caacaccaca 780
 ctactaagca caaccata caccagcccc ccctcgacgc acacggatac cccgggacca 840
 acataaaaca cagacaaca cgactccaca acaccaact aaaacgcgcc aaacaccacc 900
 cactcaccac acaagcagca gcgacaaaa caccaccca accacaaag cgcaaacacg 960
 cccccaaaa acgaccattc agagagccgg ataaaaactc aaacaagagc accacaaaca 1020
 aaagaccaca tagccaaaac acccatggaa atgtattcag caacaccacc taggactcaa 1080
 aatccccgca gcaacacaac accaaaacac ttacccccca ccaacacac aaacaaataa 1140
 taccacgaaa aaaccacaca acgcggtggc cacacccac agcaacaaa gacacacaaa 1200

53

acacaaaata ccacacacac acagaccaca ccactcaaac aacagcagtc accaacacac 1260
acgaccacac acacactacg caccagaact ccaacgcaca aaacacaaca ctactcacia 1320
caacacaacc aacgacacca caataaataa acagacaaac aaaccagacc gaacaccacc 1380
acaccaccac accagcgaca actacagaca ccacaaacaa caaaaccaa caacaaaagt 1440

<210> 78

<211> 1653

<212> DNA

<213> Homo sapien

<400> 78

ttttttattg atcagaattc aggttttatt attgagcaat gaaaacagct aaaacttaat 60
tccaagcatg tgtagttaa gtttgcaaag tgggatattg ttcacaaaac acattcaatg 120
tttaaacact atttatttga agaacaaaat atatttaaaa ttgtttgctt ctaaaaagcc 180
catttcctc caagtctaaa ctttgtaatt tgatattaag caatgaagtt attttgta 240
atctagttaa acaagcagaa tagcactagg cagaataaaa aattgcacag acgtatgcaa 300
ttttccaaga tagcattctt taaattcagt ttccagcttc caaagattgg ttgccataa 360
tagacttaaa catataatga tggctaaaaa aaataagtat acgaaaatgt aaaaaaggaa 420
atgtaagtcc actctcaatc tcataaaagg tgagagtaag gatgctaaag caaaaataat 480
gtaggttctt tttttctatt tccgtttatc atgcaatctg cttctttgat atgccttacg 540
gagttacca atttaagctt acgaggattg taagtgcaat tggctgggaa ctgacaacat 600
gtgatccaag ctactacacc cctgtgctca ctctgtcact tctcaaattc tgcgcgtg 660
aacacattca caagaacaac agggctagag cactgcaagg aaaccacaca ccaccaaact 720
caaaactcag aaacaccac atctccagag aggcacagag aggatacaaa gaatactgcg 780
caagacaaag aaatccacag actccacacc ccagggcaca acctggaacg caaaattcaa 840
aaaactaacg cggaccaaac gcggaagccc cgacccaaaa gagcacaata taaagaggcc 900
ccacgcccac gctgcgcgca ccacacgcac acccgcacgg ccaacacaca gggccctgaa 960
acacacagca ctacacacgg ggccaacacc acactactaa gcacaacca atacaccagc 1020
ccccctcga cgcacacgga taccctggga ccaacataaa acacagacaa caacgactcc 1080
acaacacca actaaaacgc gccaaacacc acccactcac ccacaagca gcagcgacaa 1140
aaacaccacc ccaaccacca aagcgcaaac acgccccca aaaacgacca ttcagagagc 1200
cggataaaaa ctcaacaag agcaccacaa acaaaagacc acatagccaa aacacccatg 1260
gaaatgtatt cagcaacacc acctaggact caaatcccc gcagcaacac aacacaaaaa 1320

54

cacttacccc ccaccaaca cacaacaaaa taataccacg aaaaaaccac acaacgcggt 1380
 ggccacaccc cacagcaaac aaagacacac aaaacacaaa ataccacaca cacacagacc 1440
 acaccactca aacaacagca gtcaccaaca cacacgacca cacacacact acgcaccaga 1500
 actccaacgc aaaaaacaca acactactca caacaacaca accaacgaca ccacaataaa 1560
 taaacagaca aacaaaccag accgaacacc accacaccac cacaccagcg acaactacag 1620
 acaccacaaa caacaaaacc aaacaacaaa agt 1653

<210> 79
 <211> 300
 <212> DNA
 <213> Homo sapien

<400> 79
 gataatcata tagcgatggt ggctctaatac atgctcgagc ggcgcatgtg atgatcgtgc 60
 gcggcgaggt acatacactt atgcacttgg aactgtactg tatcatagct acaacctctg 120
 acacaagctt tttttttttt tttttttttt ttccctattg taattgatcc attttttttt 180
 tgatcaatac aaaaaaattt ccctatttta ataaacccaa aaccttggtt atcatgggtca 240
 tactgttccc tgggttgaaa tggttatccg ttcaaaattt ccacaaaaaa tacaaaaaac 300

<210> 80
 <211> 486
 <212> DNA
 <213> Homo sapien

<400> 80
 ttactaaga tcctgcattt tatattgtta ttgttgcaaa aagaactcaa tacaaagcca 60
 atataaaaaa atcaatactc attttaaaac ataaacagta atttctgaat gtctaacatt 120
 ctctatgca aagactggga gaaagaggaa gggggagaga gaaaataaat tctttatttt 180
 aaacctttct tcacctgct gggaatgcac atgcccagc aaatgattcc agcttaaccc 240
 cttctggact ggtcattgaa gatagggttg gaagaacagt attttagaat ggcgatgaac 300
 agtgtcatta ttaactatat gtacatacac ttatggcact tggaactgta ctgtatccat 360
 gacgtagtaa cctctgacac aagctttttt tttttttttt ttttttttcc ctattgtaat 420
 tgatccattt tttttttgat caatacaaaa aaatttcctt attttaataa acccaaaacc 480
 ttggtt 486

<210> 81
 <211> 736
 <212> DNA
 <213> Homo sapien

<400> 81
aaggttctag tgattgctga ggagccggtg agcaccacag caggaggcag aaaactgaaa 60
agggcagggc tgaccagtac aggtcctgac agaggacgag aaaaggagag ctccaagact 120
tggctgcaaa tggactttgg aacgtacaga agatagctgg aggaaattca gccagaagtg 180
ggctgtgctg ttcacttggc agcggctcggc gcactgtcta agcaagcagc cagtcaccat 240
gatcttgttt attcaccact ttcactgaga aggacaccag tttatcgtaa cccaatgggc 300
gagaataagt aggaagcgtt acgtaattca gttaaacttg tcttggacga caaatttgga 360
gacttggctct tctagatttc ctgtccagca gatgctattg gaaagatgtg aattgcactg 420
agcttgtagc actattcctt ttctgcaaag atagaccata gttaacagtg cgtagtgac 480
acatgactag tgctaccgt ctttggaagc caacttggc cgtcagtcaa gtttgggcaa 540
atctaaagtt agcaaaggat ttctgccctt gaaggcacc ataactgaga aaaaacaaga 600
gaataccact cggaacacag accatataaa gtccgggggtg aggaagacac agcgggggag 660
aaggaagtgc gttccacaca cgtgggggaa gcctatttag agatcccccg tcagaggaaa 720
caggatcgca aaagac 736

<210> 82
<211> 191
<212> DNA
<213> Homo sapien

<400> 82
ctggcgtgac atctactggt catatgctgt ttcctgtgt gaaacttgtc tactccgctc 60
actaatatcc agcacaatca aaggcgagcc aggccatgtg tcccttgaca cagttctaag 120
ataaactctt ggtatctctt aacttctagg tggaagacat atacatacag ccattccca 180
tgagagggac c 191

<210> 83
<211> 200
<212> DNA
<213> Homo sapien

<400> 83
tgaaaaatctt aatcgatcac ctataggggc gatgggtctc taatctgtcg agcggcgagg 60
tgtgatggat gcggcgcccg gcggtctag ttgagagagc tgtttgcctt gttctagaat 120
tcttatcttt catttctttt ctttcttgta attcttattt ttggtttgcc tggactgttt 180
tgcatactcc aatctttctt 200

56

<210> 84
<211> 292
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (173)..(173)
<223> a, c, g or t

<400> 84
tttttttttt ttttttttgg gaactaaaaa agaacttatt aatggagggc aaggggatgc 60
aacaatacaa aaatcaaaag ctgggtgtat cagtggctca taggcgtgtt ccccggggtg 120
gtgaaattgg tcttactccg cctcacaatt cccacacaac attacgagca agntggggca 180
aacgcgaacg aggaggggaca caagagagca gcagacgaga cgaaaaaaga aaccaatgaa 240
gcggaaagga gaagaaacag aggaagaaag ggaggaagat aaacaagaaa gg 292

<210> 85
<211> 437
<212> DNA
<213> Homo sapien

<400> 85
gcgtggtccg gccgaggtcc cccccccctt tttttttttt tttttttttt ttctgtggga 60
agggctaatt ttaattaatt ttctgtaagc cttagggtta aacacacctta ggcggaatt 120
ttaactattc aaaaaaaagc agttcctacc aattccatgg gtttttaata cctctaacca 180
gatgtgggaa acgcatttaa ctggaaagca aaatatttag agagaaaata cgactattta 240
tccaaattat ataaatgct tgtacgatag gagaataaat gttgctttcc aaggaacag 300
gcacaacact tatttttata gacggcatgt taaaacgctg ggcgtacatc tatgtgcat 360
acgcttggtc tcttggttgt ggacaatggt gtatccccgc cccacattcc cccacaact 420
taccgaaca acacgat 437

<210> 86
<211> 762
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (450)..(450)
<223> a, c, g or t

<220>
<221> misc_feature

<222> (544)..(544)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (548)..(548)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (631)..(631)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (633)..(633)
<223> a, c, g or t

<400> 86
gcgtgggcgc ggccgaggtc cctttttttt tttttttttt tttttttttt ttttttttgg 60
gattttttct ttggccccc ttttttat tccccctctt ggaatttcac aaaggtaa 120
taaagaagat atttgtaa taacacagag aatttatctc acaccattat aaaattctat 180
ttctcacaca agggggataa acaaagaaca gggagtgcac cgccaaggct cagagagacc 240
tttttaaaat aaagagtggg ggcaaaatac ccccgctgcg aacacagaga tctcttgtgt 300
gggccacgtg tgaatatctc aatatcacca cgagacagag acaccacact cgtgtgtgtc 360
cccgtgaga atattatata caacactcac cactctctat ctcttatata tatagagagg 420
ccgcgcgtga tagagagtgc gtgtgtctn ccctctctag agagatctct ctctatatct 480
ctctatagag agagaggctc ctctctgga gagatatctc tcctctctta tatataagag 540
cgcntggngc gcgtatatct ctgcgtggc gccacatacg cgtgtgtgtc tctcgcgtgt 600
gtgtgtagaa catgtgtgtg tatatctcgc ngctctcac acatatctct ctccacacaca 660
caacacattt tccgcgagca ccacaacaac taacgtggca cccacacaa cccaacaccc 720
caaccacac acccacacca caagcgaca caaccacaac ac 762

<210> 87
<211> 476
<212> DNA
<213> Homo sapien

<400> 87
gggattta atatacgat ggtcttaatc attcgagcgc gcagtgtatg atcgtgtcgc 60
ggcgagggtt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 120

```

tcattttttt tttttctttt tttttttttt acaaaaaaat tttctttttt tataaccccc 180
aaatttgctt ttttggtttt ttttggtaaa tttttttccg tccccaaatc cccacaacaa 240
tcatcaacaa aacatgtcat ggtagatgca gtcccgctg ctcaccagca cactacgctg 300
tacagctacc aacacgagct cagagagcag gacgaagtac atgcaatcgt agctgactag 360
agagcactga catgagcgga gtggacgata tcacggtcgc agagcgtagt aaagtcggca 420
agtgagctga aggacatagg agatagatca gatagtagca cattggctcat atacgt 476

```

```

<210> 88
<211> 842
<212> DNA
<213> Homo sapien

```

```

<400> 88
gccggcccg gccggtacac tgtccaacaa gtataatgcc ttgagaagtc tcgtatttca 60
aatcctcttg ggatcgcgca tggcgtagc tgtggtgctg acataaagta cgtggtgctg 120
aactacgtgg agttcttcta gtcttgggta ctagtgcgga ctataccact ggatcaggtc 180
ttcgatcttt agttcgtggg aacatatgtt aacgagccaa gctacgaaga catggctcgc 240
cagacttgtg ggcaacgcac ggggtgcagg ttgtcagtgc ttattgggcg tgtgtaagta 300
caagcgcaat tcgtagcccg catagacatg caaggacatg gactagaact tgcccaagat 360
gcctacaacg aagagcgagc gtgttaacaa actacgcaat atgcaatgac tatggcctca 420
gtagagtaat attgagtagt gcctccatgg gttctagttt aagggcgata acacctagtg 480
tttgaatttc acacattctt aaacagtact aacgttttag agacctaggg tacattcttg 540
catggacatg ggtagcgtat ctaaccctag aaataagaac cacgtcactg aagaatagac 600
ctacttccaa ggtaacccat cgttttttag aaaacccgag gatttaaccg cgagagagaa 660
tcctaggagt ctcaaggaag agtttaactt aaagggggtg ggctccgtgg gaaaggggtg 720
gtttccccta aacgaattaa tctcagagtt attcccgtgt ttaaatttaa caagtcttcc 780
cattttaagc caagttggca aaaaacacca aaaacaaaca aaaacaaaca caaaaaaaca 840
gt 842

```

```

<210> 89
<211> 1729
<212> DNA
<213> Homo sapien

```

```

<400> 89
acagaattcg gcacgagaga ctataccact cccataccct ataactttgt ttgttctatt 60

```

59

tcacacatat aattttccga gacaagatgt tctcatTTaa gcaacaagaa gattcgtctc 120
tcgctattac tgtaactgct gtttatatcg tcatgtcccc gaaaggTccc tgtcttccct 180
gaatggtctc taccaacttc acctccggtt ctaggtgtca tggctgcccc aagagtctag 240
agacgacaac ttctccgctt cctcggcgat ggcggcgTcc gggagcggta tgtcccagaa 300
aacctgggaa ctggccaaca acatgcagga agctcagagt atcgatgaaa tctacaaata 360
cgacaagaaa cagcagcaag aaatcctggc ggcgaagccc tggactaagg atcaccatta 420
ctttaagtac tgcaaaatct cagcattggc tctgctgaag atggtgatgc atgccagatc 480
gggaggcaac ttggaagtga tgggtctgat gctaggaaaag gtggatggTg aaaccatgat 540
cattatggac agttttgctt gcctgtggca gggcactgaa acccgagtaa atgctcaggc 600
tgctgcatat gaatacatgg ctgcatacat agaaaatgca aaacaggTtg gccgccttga 660
aaatgcaatc gggTggtatc atagccaccc tggctatggc tgctggcttt ctgggattga 720
tgttagtact cagatgctca atcagcagtt ccaggaacca tttgtagcag tggtgattga 780
tccaacaaga acaatatccg caggggaaag tgaatcttgg cgcttttagg acatacccca 840
aagggtaca aacctcctga tgaaggacct tctgagtacc agactattcc acttaataaa 900
atagaagatt tgggtgtacac tgcaacaat attatgcctt agaagtctca tatttcaaat 960
cctctttgga tcgcaaattg cttgagctgt tgtggaataa atactgggtg aatacgTtga 1020
gttcttctag cttgcttact aatgcagact ataccactgg tcaggTcttt gatctttagt 1080
tcgtgggaac atatgttaac gagccaagct acgaagacat ggctcgccag acttgtgggc 1140
aacgcacggg tgcaggTttg tcagtgttta ttgggcgtgt gtaagtacaa gcgcaattcg 1200
tagcccgcat agacatgcaa ggacatggac tagaactTgc ccaagatgcc tacaacgaag 1260
agcgagcgtg ttaacaaact acgcaatatg caatgactat ggcctcagta gagtaatatt 1320
gagtagtgcc tccatgggtt ctagtTtaag ggcgataaca cctagtgttt gaatttcaca 1380
cattcttaaa cagtactaac gttttagaga cctagggTac attcttgcTat ggacatgggt 1440
agcgtatcta accctagaaa taagaaccac gTcactgaag aatagaccta cttccaaggT 1500
aaccatcgt tttttagaaa acccgaggat ttaaccgca gagagaatcc taggagtctc 1560
aaggaagagt ttaacttaaa gggggTgggc tccgtgggaa aggggtggtt tcccctaaac 1620
gaattaatct cagagTtatt cccgtgttta aatttaacaa gtcttcccat tttagccaa 1680
gttgGcaaaa aacacaaaaa acaaacaaaa acaaacacaa aaaaacagt 1720

<210> 90

<211> 1378

<212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (547)..(547)
 <223> a, c, g or t

<400> 90
 gcggggccgcc cggggcgggta cccggggccca ggaggacgcc gagcgggcag ccccgtaggc 60
 cgcgtggacg tctagcgtg ccgtaggatg atgcgctccg ctacgagaag gagctggtcg 120
 aagtcgctgg aaggaagtct cgtaatgaag tctcaaggat gtgggacgga tgttgcccgt 180
 tctattgacg aaggaagtat ggccagagtc cccaggtgtg tgacgccagg tggagccagt 240
 ggtgcatcga gaggcaatag gggcagagga gtcggggcaa agctgctgtg acatggtcgg 300
 ccgatgggag gccatccttg cggatttcgc ttctcttcgc tgaagatgct tagtgatagg 360
 gggccggctg caatcgcatc tcataccgag tacgattccc agcattcgca cagtcagtag 420
 cgtttagccg cgctatggac gggacgcgag agccgtgtcc gtcccccttg gggaggtctc 480
 tgggcgtgtg aagggcgaaa gctaggagta acaagggtgt atgataggta tatgtgcccc 540
 tacttgnaga gggggggacc aatggggggc cttctaattg tcgcgctggc cgcttgttgc 600
 atgatacatc tacagcagcg gaaaccggca ttttgacgaa attcactcag acataaaata 660
 caataaacgg gaaagtcgaa cacgcggcag taaaacctgc ccacgcgcgg ggcgagacca 720
 atctccaggg gggcccaaca taagtgagga gcgtagcccc taggggcgga gtggatagag 780
 ggcgcacgac ggcgcccggc aaaccaccca aatttcacac gcggggagaac agataaccgg 840
 aggcgacaca caaaccccta gagaataaag ggcgacaggg gggcaaagag cgaggatacc 900
 caggaggaag aggaaactac aggcacacac caagagagga aaaagtgaac atacacgagg 960
 gacgtggcac atggccaaaa aatggacaac gggacggacc cacatatcca taagatgcgg 1020
 ctcggatgac cacacacaga cccaacatgc ggtaaagacg acaacacacg ggaccggaac 1080
 catgactaag gaaagccacc accgagacaa acagcaaaca caacctatat tcacaccgtg 1140
 ggtacacaca gtataggaca aaagaaatcc actacaaca tatgggtagg agcaccacat 1200
 agagtgacaa cgaaagggga ttgggatcac aaaccagac aacaacgtct gagaagcaca 1260
 tacaggaccc cacacgagag acgcaccaca acaccgacgg aacgtgccct gcaaaggat 1320
 agaaccacgg cggatacaac cggaactcac accatcacca acgccacaca acaaaaat 1378

<210> 91
 <211> 1278

<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (827)..(827)
<223> a, c, g or t

<400> 91
gcggccgccc gggcaggtcc cccccctttt tttttttttt tttttttttt tttttgtggt 60
ttaaaaaaag tggacttttg cttttttcct agtgtgggag aggggtggggc ccgtgcagtg 120
agaggggtggt aaggtgtgcc ccaggggaag tgggggtgggc ccgtgtgtgg aaaataatga 180
aacaaaagag ggtcgtggaa agagaaagag ggggggtgggt ttgtaagagg cccagtgcgc 240
acaaggtgtg cgcgctcctc agggctcgcc gctctctctg tggaggagtg gagccaccgc 300
tctgtgtgag agagaagaga gagtctcagt gtgcgcgggc gccatatat gctgtgtgag 360
cgaccacaa atctcaatat tataaaaaat cttcgtgtgg agacaatctc tatagcgcgt 420
gtccccactc tccggtgtgt gtgtgtgtgt gtctcccagc acttctcttc tcaacacaag 480
agcgcggctc gagagtgaag cccccgggt gggctctcct ctgctgcgtg ggggtccctg 540
tgtgtggaga gagggcagca gacaataaga ttcgctgtgt gtaattctcg atgatgaaag 600
ccccgtgag cggcgataa acacctgcgc gtggcgcca aatgaagcg ctgtgtgtcc 660
cgcgcggtg gtgtgtgcac acatgtgtgt gggtatctcg gggggtcta cacaaaaatg 720
tctccacac accacaactt atttgtggcg cgcagccac acaaagcctc cacaacgcgt 780
cggggttgtt cttcctgag ccaccaaca acgactgatg cgggcnaca aacgaatcag 840
cctaaaacac aacaaaaacc acagcaccac aactgaccg tagacacca ccaatagac 900
aaccacaaac aaaaacaca cacacaacac accaaccaac gaatcacaa caaaccacaa 960
caaaacacag acaacaacc aacagcgaga caagagaacg gagataggca taaggcgaga 1020
tgaacctag agctctcgta gagaacctg cacaccaact agcgaataac cgaccgcaat 1080
ggtcacctag taaaagccg accagaataa ccgatatgag atccaacca cacaaaaacg 1140
aaagatgata aagatgaca acgtaaaaat caaataatga gatagacata agacgaaatt 1200
gaagacaaca gactcggat gagaatacag aaacaatcaa gagtagcaaa gagacagaac 1260
aaacacaaag gaaccccc 1278

<210> 92
<211> 421
<212> DNA
<213> Homo sapien

<400> 92
 cgcccgcccg ggccagggtt tttttttttt tttttttttt tttttgggga aaaaattttc 60
 taaaagggct cttttttttt ttttggggac ggtgttaaaa aaaataccgc ggtgttttta 120
 aaattatttg cattatgagt tcccatttaa caaatcgtg tgtgttccca aaatatagtt 180
 ctctttttac ccagggtctc gtgggttaaaa tataccaaca cccgggtata aaattttctc 240
 tgtgggaatc cttattccat aaaaaatggg ccccagggtt tctcacacca ctagtgtgga 300
 aaatgttgtg ggggtgagat gagaaatctc actttttatt atatctcaac gccggggggg 360
 aaacctcgtg ggccaatagc cgtgttcccg tgggtggaaa gtggttatcc ccgccccaaa 420
 t 421

<210> 93
 <211> 544
 <212> DNA
 <213> Homo sapien

<400> 93
 acaaattctta agtataaatt tactgagttc ttgcagacat atacacctgt gtaaccaaac 60
 ctttccaaaa tttagaccat tgcaatcatc ttcagaaagt ttcttaaate cctctcccag 120
 tctatcccct cccaccccct cagggtataac tactgttctc attcttttat atcaaagggt 180
 aaatttacct gttctaaact tcatatgagt gaaattatac aaaatgtaat gcttctttca 240
 cttagcataa tgtttttgag atttattcat gttgttgcac gtgtcagtga ttcattttctt 300
 tttattgctg agtattcttt cgtatgaata tatcacagt tgttttttta tctttctggt 360
 gatggacact gggctctttc tgcttgtttt ttactgttat gaataaagct gctatgaaca 420
 ttcttagaca aaaaaaaaca acaacaaaac aaaaaagtcg gggggaaaacc ggggaaaaag 480
 gggacccggg ggggaattgg ttccccggcc aaattcccc caattttgcy acacaaaaga 540
 caac 544

<210> 94
 <211> 5631
 <212> DNA
 <213> Homo sapien

<400> 94
 gcggccgagc ctctgcggg gatgcggtt taacgcgctg gggcctccct ttatagggtc 60
 acatgtgatg gggaggagga tacaagcccg gaatctggga taggtggtaa tagggccacc 120
 agctgctttt agttttacgc ttcttatgtc ttacctggaa acaattattc taaaagttgg 180
 tcacctcatt aagccaaaac atgagggcag agtggaggaa ggatgcaaaa agttttaaaa 240

taatatagga aagtactaaa tttagaatg tctggacaac tcattccagg tcacctatag	300
cctatgagag aggaagaata tttttgaca attatcggcg ctgtgtcagc agtggtgcat	360
ctgagccaag aaaactttat gaaatgcca aatgttccaa atcagaaaaa atagaggatg	420
ctttattatg ggaatgcccc gtgggagata tacttcccaa ttcacagat tataagtcct	480
cactcatagc actgactgct cataattggc tacttcgtat atcagcaact acgggaaaaa	540
tccttgagaa aatatatctt gcaccttatt gcaaattcag atacttgagc tgggacactc	600
ctcaagaagt cattgcagtt aagtcagctc agaacagagg ctgagcagtg gcccggcagg	660
caggcattca acaacatggt ttgctgtacc ttgcagtggt ccgagttcta cctttttcac	720
ttgtagggat tctagagatc aacaaaaaga tttttgggaa cgttacagat gctaccttgt	780
ctcatggaat actgatttg atgtacagct caggactggg cagactctat agcttccaaa	840
ccatcgctga acagacatgc caccactgct ctttgagggtg tcacccctgg agaatgcttt	900
tcagattgga ggccatcctt ggcactacat cgtcacacct aataagaaga aacagaaagg	960
agttttccat atttgtgcc taaaagacaa ttccctggca aaaaatggga tccaagaaat	1020
ggattgttgt tctctagaat ctgactggat ctatttccat cctgatgctt ctggtagaat	1080
aatacatggt gtgccaaatc aagtcaaagt ttggaagcta actgaaatag aaaataatag	1140
ttctcagcat cagatctctg aagattttgt cttttggcc aacagggaga accataaaaa	1200
tgaaaatgta ctactgtta cagcttctgg acgggtggta aaaaaagtt ttaaccttct	1260
ggatgatgac ccagaacaag agactttcaa aattgtggac tatgaagatg agttagattt	1320
gctttctgtg gttagctgtta ctcaaataga tgctgaagga aaagctcacc tggatttcca	1380
ctgtaatgaa tatggaactt tacttaaaag cattccacta gtggagtcac gggatgtgac	1440
atatagccat gaagtctact ttgacagaga cttggtgcta cacatagagc agaaacccaa	1500
cagagtcttc agctgctatg ttaccagat gatatgtgac actggggaag aagaagaaac	1560
cataaacaga agctgttaaa aagagtgaga taattgtaac ctaagagact tttagccaaa	1620
caccccagca gctgcgtcca atccatttta ttatctgcat ggcacattct ccagtatttt	1680
ccaaaaaagt cttgtgttga cttcagatga ctatgacttc ttttttaaac tcttctgtga	1740
aaagatgggtg aggacttcat tttttttaa ggtttttttag aatactgttc caagaagttt	1800
agtgttttgc agctttgagc taggtggtaa tgcaaatata aaatgctggg aacagaaaag	1860
gacagggttaa ttccaattgt tgaggagtta agtcattgat ggggtgggtc attgatgagt	1920
tcttaaagga tggatggaa ttttgttgt taaggctagg aaagacagg agagacaaaa	1980

gtaaacaatgc agaaagaaat cttatatcct ctataccaaa ctttgcttaa ggatgagaaa 2040
tgagatgtgt tatgtgagaa cattatcttg agcccaaat gtgtcatcac agtttttaaa 2100
aatcttatat atgtatttat atgtgttcg tattgtata tagtatcagg aattggttct 2160
agttcccaaa ttatcttttc ttccttggtt ttgttctctt ggcttgatgt tcacattgaa 2220
tatttgtgtt tctatatagg ctaatgtaaa agattccaag caaaccttaa gtgaaattgt 2280
tttctgattt gcatcctgtt tagctcttaa tgtatctaag gatgttctca tctcaccatt 2340
ctactcattt agtgagtttt ctgatcttgt ttaggcaata ttgcatact tatgcaataa 2400
gataaaggta cccttgctg cagtagttct gtttctgta gaaaagtga taaagagtc 2460
cagaagaagt tcttactagc ttgggagtta cctgattaac cagagaaaat ttttggtta 2520
cttatggaac aagcattatt tcttcttctg taggaaagat ctaaataatgg tccttgactt 2580
ttaataatca ttctttagaa tgtaaataa aggcaacca agtaaaggga gaaaatgttt 2640
cttgtgtctt cctgttgag aaattcagtt gcttccattt cgcagtgtct gcacatttat 2700
ccgatgtaac ctcaaagaa taactggtta taagggaagg aaacagcagc aacaatcatt 2760
gctgattcaa gttaagggtt aaaatatgga attttagct tggatgattt atattaaaat 2820
cttccattt tttttttca gttttggtt gatgccatgt taagaatgat gtgaattctt 2880
cccagttctg ccctgggtgct agacattgcc ccatactttc aattagacac tagctgtatc 2940
taaatagtcc cactcagtaa acttacatct tgaaaaacaa gaccagtaag aggccagtga 3000
aagtactaaa gaaagaaacc aatgttgtgt gagtttcaaa gcagctgcaa tgctgtgtaa 3060
aagtagagtg ttcattctcc atttccaaga gtgtttcaga ataggatgtc ttaagacttc 3120
agtcagtca gagatctttt ttttaggtg attattgagt ttctccttct cctttaagtc 3180
atcaccttc ttttatgaa tgatagtaag gaactcgtct attctgaaag gcatttgaga 3240
aatagctgaa ttcctggctg cttttttgct gggggtagat ggtggaatac ttctggtcta 3300
gatataactt accactaaga aacccccagt atgtcaccac tgcctaaatc taactagacc 3360
agggtccaaa tgccatccag gccaggcagg aaatatacct catgtgaaag acagtaagga 3420
gttggtggga gtgtaacaaa caggagagct atgccccaac taaaaggagc agctgctact 3480
gcttagtttc agccagttgc aacagtatgt gggaatgtag gctgcatggt tgttaacaag 3540
atagatggta aaaagatgcc agaagataca gaagatagca aagaatgtgg ggaatttgga 3600
taccacacat agcgagagac aatgaagcat gcttccagc tcgccagagt gtcacacagc 3660
tgctcattct gccacctgcc agacattaat gtcttcctgc cctacctaaa cccctcttt 3720
acctgatatt ttaattcgag actctagcta catgcccacc tacttaacag gtactagtga 3780

cagggtacaaa acattatggg taacaattct gagtgtttta tgcaagccca ggtgaagcag 3840
ggtagcttcc atcagcaggt acagacgtta cgctgaaaag aggtgcattc tgcattgcac 3900
tcctggatct aagtttctgt attctcagag catcaatgca gcaagcttat tgttctctaa 3960
ttttttacaa ttttatcac aactctggga gaaaacaaaa caaacctat cctattttact 4020
atltgtgcta cctagtgagg agataccgct ctgttttagac aaattaaggc acttcacatt 4080
cttccaccaa ttgaaagttt tgtatcttac agttcttttt ttaaataata tattttattga 4140
gcactttcta tctactagtc actgtgatac agtataagta aagtgggttg tctcatTTaa 4200
tattcagaat aaccacatga agtatgaact gccattatct tcccccttg taaaaatgag 4260
gaaagtgagg ctcacagaag ttaattggcc caggggtccca caactagtca gtgcagaggt 4320
gggaaacata accagatttg ttcggcatga acttgtgcca aatttctctc aaagctctca 4380
aaaggcaagg catgttattt tatcccaatt tagcatacca acaactataa tactagatat 4440
gtaggaaagt gcttaataat cgttttttac tgatgattca gtgtctaaat tttgaacaaa 4500
tttgggtaag atacaagtca cacataaatt gacagaaaat gtagttcttc attcaatggt 4560
tagcagtcac taaaaggtag tttcccttg tttgtggtga taatcagtat tagtagtttt 4620
catattattt ggcttcata ttaatcattt ttatattttc ttctccttct taccatgttt 4680
acttatatca tccatctttt agaatcccag ggagctaatt tctgggtccct gtgttgctat 4740
caaatctgta tcttgagaa agaataattt atttcaaaca agggacatac aatagaaaga 4800
taagacctac tgaggtcttt tcccatcat tttattatga aaaatgttca aacatacagt 4860
aaaattgaaa gaattttata gtaaatactg accacgggga ttctacatct tactctactt 4920
gttttattat ttctctatcc agcgtacttt ttgatggatt tcaaaataaa ttgcagttgc 4980
tgatatactt cccctagta cttcaactgc agattattaa ctagagttta gtattttatt 5040
agttttttaa tttttttgat ttaagattta cctgcaataa aatgtacaaa tcttaagtat 5100
aaatttactg agttcttgca gacatataca cctgtgtaac ccaaaccctt tccaaaattt 5160
agaccattgc aatcatcttc agaaagtttc ctaaatccct ctcccagttc atccctccc 5220
caccctcag gtataactac tgttctcatt cttttatata aaagggttaa ttacctgtt 5280
ctaaacttca tatgagtga attatacaaa atgtaatgct tctttcactt agcataatgt 5340
ttttgagatt tattcatgtt gttgcatgtg tcagtattc atttcttttt attgctgagt 5400
attctttcgt atgaatatat cacagtttgt ttttttatct ttctgttgat ggacactggg 5460
ctctttctgc ttgtttttta ctgttatgaa taaagctgct atgaacattc ttagacaaaa 5520

66

aaaaaaaaa aaaaaacaaa aaagtcgggg ggaaaccggg gaaaaagggg acccgggggg 5580
gaattggttc cccggccaaa ttcccccaa ttttgcgaca caaaagacaa c 5631

<210> 95
<211> 96
<212> DNA
<213> Homo sapien

<400> 95
aggtaggagg ggtagagga gggaggactg acgtgaggaa gaggaggacc attcggacaa 60
tgtattagga cactctcacc aagctggggg atcatg 96

<210> 96
<211> 495
<212> DNA
<213> Homo sapien

<400> 96
gcgtggtcgc ggccgagggt ggaaaaaaaa tttttttttt tttttttttt ttttggaaac 60
aaagtctaatt tctgtcggc aggttgaggt ggggggtgtg atcttggtca ttgaacctcc 120
gcctttgtgg gttaaaaaaa ttctcctgcc tcagcctcct gagtagctgg gattataggc 180
acacaccacc aaaccagtc taatcttttg gtattttttt tatagtagag acgggggttc 240
gccatgtcgg ccaggcggtg tcttgaactc cgtgacctcg gggggctcca cccgcctcgg 300
cccccaaag gtgcgggcat tacgggcgtg aaccaccggt gccgggccgg aagtctttaa 360
aaaaataagg gtgatactac atcttcaaag actgggggat aactcagggc ccatagctgg 420
ttcccggtg tgaacatttt tactccgcct cacaatcccc cacaaatact cgaaaaatgc 480
ggaaaaaaaa aaaga 495

<210> 97
<211> 1374
<212> DNA
<213> Homo sapien

<400> 97
cggccgggca ggtaccatgg caatgaagcc ggatgtctgc cttcctgca accagctccc 60
tcaggagtgt tagcttcaag ggtgactggc cggatgagca tgctggagag catcagtggg 120
gaggcgatgt gggaaccgg tctgcctgac ctttctattt ctgaatccag acccatcgtg 180
caccatggct acgtgtacca atctggtaaa acttctcatt gtcactcca gtgaaacgcc 240
cagtagtatc acgaacacat ttccgactcc cactgtagtc atagaaatcc gtggcatcca 300
actatagaag tgggtgtcta caacgtgata catcgtaggt ttaattaaat agaaataaaa 360

67

taaaacaaga caaagacaga taacaggcat gagggagaga aagccgtgcg ggagccaaac 420
 aagacgagat taaggcccga ggggagagca aacagtgagg gtcacacgcc cgggcgccaa 480
 caagagtacc acagagcaaa agagaagata acaaaatata aagggggaac aaatgacgaa 540
 aagagaaaaa cacaatgtag gctagaaaaa acagaggaaa aaatagcaag aaatacaaac 600
 aaaaaattat cccaaagcaa acaaaagcaa atagaagaga aggcaagaa gaggaaagaa 660
 aaaagacacc aagaaggaaa gaaaagaaag gaaaaaaga aagaaagaca caaggacaaa 720
 aaacagaaaa aaaaaaaaaag aagaaaaaaa aacagcaagg agaaaccaga acaggaagcg 780
 aagacagaag acagacgggc gaggcaaagc aagaacagaa cacagaaaag acccacacga 840
 gaaacagaga aagcgaaacg tacagacaga aggaacgaac aggaacgaga cagagacgca 900
 aaagcagaag aaagacacac aagaacagga acaaaacaga ccgaaagaca aggagccacg 960
 aacggagaaa gaaaagacga ggaagagagc gagcggagag gaagaagagc aggagagacg 1020
 agagaagaaa ccgacgagag agagagggcg gagcggcgga gaacaaaaga agacgcagcg 1080
 agccgcacac gacgaaccac gcacacggag aaacgcaacg cagcaacagc gcgcgacgac 1140
 aaacggaaga cgaaagagaa aaacaagcac acaaagcaga caacgcaaga acgacccata 1200
 cgaccgacga cagaaccaag cgcagagaca acagcaaaag aaggagacag gaaacgcgag 1260
 caacagcaac acggaacgac gacgacaaaa acgaacacac ggaggcgcaa gcgcagaacg 1320
 aaagaaacaa cgaacgagaa cacaacagaa acacacgaag cacacacaca acgg 1374

<210> 98
 <211> 1713
 <212> DNA
 <213> Homo sapien

<400> 98
 ggaacaaatg tcatgccagt gggagttaa gtgccagcat ggagaagagg agtgcaaatt 60
 caacaagggtg gaggcctgcg tgcttggatg aacttgacat ggagctagcc ttcctgacca 120
 ttgtctgcat ggaagagttt gaggacatgg agagaagtct gccactatgc ctgcagctct 180
 acgccccagg gctgtcgcca gacactatca tggagtgtgc aatggggggac cgcggcacgc 240
 agtcatgca cgccaacgcc cagcggacag atgctctcca gccaccacac gagtatgtgc 300
 cctgggtcac cgtcaatggg aaacccttgg aagatcagac ccagctcctt acccttgtct 360
 gccagttgta ccagggaag aagccggatg tctgcccttc ctcaaccagc tccctcagga 420
 gtgtttgctt caagtgatgg ccggtgagct gcggagagct catggaaggc gagtgggaac 480
 ccggctgcct gccttttttt ctgatccaga cccatcgtgc accatggcta cgtgtaccaa 540

This page is not part of
the pamphlet!

WO 02-068645

3/4

Date: 06 sep 2002

Destination: Agent

Address:

68

tctggtaaaa cttctcattg tcactcccag tgaaacgccc agtagtatca cgaacacatt	600
tccgactccc actgtagtca tagaaatccg tggcatccaa ctatagaagt ggggtgtctac	660
aacgtgatac atcgtaggtt taattaaata gaaataaaat aaaacaagac aaagacagat	720
aacaggcatg agggagagaa agccgtgagg gagccaaaca agacgagatt aaggcccgag	780
gggagagcaa acagtgaggg tcacacgccc gggcgccaac aagagtacca cagagcaaaa	840
gagaagataa caaaatacaa agggggaaca aatgacgaaa agagaaaaac acaatgtagg	900
ctagaaaaaa cagaggaaaa aatagcaaga aatacaaaaca aaaaattatc ccaaagcaaa	960
caaaagcaaa tagaagagaa ggcaaagaag aggaaagaaa aaagacacca agaaggaaag	1020
aaaagaaagg aaaaaaagaa agaaagacac aaggacaaaa aacagaaaaa aaaaaaaga	1080
agaaaaaaa acagcaagga gaaaccagaa caggaaagcga agacagaaga cagacgggag	1140
aggcaaagca agaacagaac acagaaaaga cccacacgag aaacagagaa agcgaaacgt	1200
acagacagaa ggaacgaaca ggaacgagac agagacgcaa aagcagaaga aagacacaca	1260
agaacaggaa caaacagac cgaaagacaa ggagccacga acggagaaag aaaagacgag	1320
gaagagagcg agcggagagg aagaagagca ggagagacga gagaagaaac cgacgagaga	1380
gagagggcgg agcggcggag aacaaaagaa gacgcagcga gccgcacacg acgaaccacg	1440
cacacggaga aacgcaacgc agcaacagcg cgcgacgaca aacggaagac gaaagagaaa	1500
aacaagcaca caaagcagac aacgcaagaa cgaccatac gaccgacgac agaaccaagc	1560
gcagagacaa cagcaaaaga aggagacagg aaacgcgagc aacagcaaca cggaacgacg	1620
acgacaaaaa cgaacacacg gaggcgcaag cgagaaacga aagaaacaac gaacgagaac	1680
acaacagaaa cacacgaagc acacacacaa cgg	1713

<210> 99

<211> 1448

<212> DNA

<213> Homo sapien

<400> 99

tggtcgcggc cgagcgtagt tttttcccaa acatgtgtgt atatttcgta agcagttaga	60
acataattact agacttgact ctgacgaact tcaccctctg aaaattcctt gacaccactt	120
cctaacttta catacgtgct catggcttac acataaacat ctactaaaga cggcacttct	180
ctatcctcta tactgcaacg cctaacctcc agattccgac tctagcgcta acctaacgac	240
tcaatacctt gctccatacc ttgctcctct tgcttcctca ctttcctcta attctcttca	300
tattctctta acacaacctc aagagtacta ttctcttaac ggcacacgaa cgctaacgag	360

69

cacagcatct gccttgccac gaaaatgcct tcagacagaa tgcattctctt catcttaaaa 420
 atggcttccc ttaggcaccc cacgggacaa ccttgcaagc tcaaattctca gggcgctcac 480
 tgcacacaac tctcccacgc tctcactacg gcttctctac aactccttac tctgggctac 540
 aactcttcaa acattaacgg cttttctctt caacattgca ccttacaaaa cattgaacaa 600
 ggcttctctc tctagaaccc acaacaacac actacacaca cacacatacc acaccacacc 660
 acacactaga caagacagcc gactactcgc tgcggagccg gaacaacact cctcatacag 720
 acgcgcgcac catacacgtc ggctgtgtga tcaccacccc aaggcgcggt gtgcagcacc 780
 acaccgtcgc ggggatcatc acatcacggg gacatcacca acagaagaag attcccgccc 840
 acagagaaca aaacagtcta ggtgcacaag tcaaaaagat gtagggctcg tactacgtaag 900
 catcgatagt gctcgcgacg tagaggttac cgagtgtctg gcacagcgga tgggtgacagt 960
 gagtgataaa tgattaggat ggcgccacgg tcgacaagat tgtgttgatg gcgtgctcgc 1020
 gtgttcgtga ctatctatcg tttgatgtga tctgccagtt gactattatt agtactttcg 1080
 cgatgagtgc gggcgctcgc gatggacgtg cgccgtagcc gacgcggagc agctgagtgc 1140
 agaggcgctg ctgagccaag taaataatgc aaggggcaag gtcgggcgga aggcgcgggtg 1200
 cgcggtgggg aagagtgagc agaggtgacg aggcggagag gggaccgacc tgtgatggga 1260
 gggcgagcgg gaggtaggag gaagcatgga ccagtagtag atgtgcgagg agaggtgtgg 1320
 tgagcgagca gaaggaagcg cgacacaaaa gtgcgagagg acggagacga ggacagatat 1380
 ggacgcctgg agagaggagt caacgagagg gcacgctaag cggcgagagc ggtcgaggcg 1440
 aacgaaac 1448

<210> 100
 <211> 1786
 <212> DNA
 <213> Homo sapien

<400> 100
 atttaataga ctatataggg atttgctgc ggcaagaatt cggcacgagg gatgccaaag 60
 aagaccatga aagaacacat caaatggtct tactgagaaa gctttgtctg ccaatgttgt 120
 gttttctgct tcatacgata ttgcacagta ctggtcagta tcaggaatgc ctacagttag 180
 cagatatggg atcctctgag ggccacaaac tgtacctggt aagttctaga gcctttagt 240
 tttaaatttt aatgatttga tatgctctgt agtaatatta attttgtaa gatgtgttta 300
 catttgtaat tgctcttga ttttcttaaa gtaatagtc agttttaatg ttttaatgtt 360
 tgtacttttt tcccaaatg tgtgtatatt tgtaagcagt tagaaatata tagactgact 420

70

```

ctgagaattc accctctgaa aattccttga caccacttcc taactttaca tacgtgctca 480
tggcttacac ataaacatct actaaagacg gcacttctct atcctctata ctgcaacgcc 540
taacctccag attccgactc tagcgctaac ctaacgtctc aataccttgc tccatacctt 600
gtcctctctg cttcctcact ttcctctaata tctcttcata ttctcttaac acaacctcaa 660
gagtactatt ctcttaacgg cacacgaacg ctaacgcgca cagcatctgc cttgccacga 720
aatgccttc agacagaatg catctcttca tcttaaaaat ggcttccctt aggcacccca 780
cgggacaacc ttgcaagctc aaatctcagg gcgctcactg cacacaactc tcccacgctc 840
tcactacggc ttctctacaa ctcttactc tgggctacaa ctcttcaaac attaacggct 900
tttctcttca acattgcacc ttacaaaaca ttgaacaagg cttctctctc tagaaccac 960
aacaacacac tacacacaca cacataccac accacaccac aactagaca agacagccga 1020
ctactcgctg cggagccgga acaacactcc tcatacagac gcgcgcacca tacacgtcgg 1080
cgtgtgtatc accaccccaa ggcgcggtgt gcagcaccac accgtcgcgg ggatcatcac 1140
atcacgggga catcaccaac agaagaagat tccgcgccac agagaacaaa acagtctagg 1200
tgcacaagtc aaaaagatgt agggtcgtta cacgtaagca tcgatatgtc tcgcgacgta 1260
gagggttaccg agtgctgggc acagcggatg gtgacagtga gtgtataatg attaggatgg 1320
cgccacggtc gacaagattg tgttgatggc gtgctcgcgt gttcgtgact atctatcggt 1380
tgatgtgatc tgccagttga ctattattag tactttcgcg atgagtgcgg gcgtcgcgga 1440
tggacgtgcg ccgtagccga cgcggagcag ctgagtgcag aggcgcgtct gagccaagta 1500
aataatgcaa ggggcaaggt cgggcggaag gcgcggtgcg cgggtgggaa gagtgagcag 1560
aggtgacgag gcggagaggg gaccgacctg tgatgggagg gcgagcggga ggtaggagga 1620
agcatggacc agtagtagat gtgcgaggag aggtgtggtg agcagcaga aggaagcgcg 1680
acacaaaagt gcgagaggac ggagacgagg acagatatgg acgcctggag agaggagtca 1740
acgagagggc acgctaagcg gcgagagcgg tcgaggcgaa cgaaac 1786

```

```

<210> 101
<211> 467
<212> DNA
<213> Homo sapien

```

```

<400> 101
taaacctgag caaggcgtga gtggggcaaa catctgggga tggatagata gagatggcta 60
aactcaaadc aattagctga gagatcactg cccgatgtac agtctgggag cctcgtgtgc 120
ctgatgaaat atgatcggca acgcgcagag aaggaaaggt gcgtataggg agctctaacg 180

```

71

aaatatccaa tcaagaatcc acgagagaaac tacaacatag aagaataaaa gaaaggaaaa 240
 aagagataga agagaaaaaa aagaaaaaac aaacatataa aaataaaaat cgagcagaga 300
 aaaaacagag aaaaccaaca aaaagatcaa gaaagagaaac aagacaagaa aaaagaacaa 360
 ttgtagaaac aaaaaggcaa aagaaaaaag accaagagaa aaagaaaaaa aaggaccagg 420
 agtattagaa agaaagataa aaacaacaaa ggacaccaga cacaact 467

<210> 102
 <211> 103
 <212> DNA
 <213> Homo sapien

<400> 102
 tcccgactc attcggtgag tcaaccggta gacacgagtt atacgcactg gaggagcatt 60
 gagaataatg gacaagtgtg ccgggtagag tatcggtccat caa 103

<210> 103
 <211> 724
 <212> DNA
 <213> Homo sapien

<400> 103
 gagcgggcgc cgggcaggtg gagacaggtc tctctctctt gcctagctgg gagtgcagtg 60
 gagtgatcat agctcactga ggcttgatac tcctgggctc gagcaatcca cctcagcctc 120
 cagaagtagg ggagactaca tgatgtgtgc caccatactc agctaatttt taaacttttcg 180
 tatagacagg gtctccctgt gtagcccagg ctggcctcga actcctgacc tcaaaaaatc 240
 ttcttgctt ggctcccaa agcactggga ttataggtgt gagccatggc gcctgggtcat 300
 aaattctatg ttatttgttg tttgttcgtt tattagagat ggaatctctc tctcttgacc 360
 aggctagagg gctgtggtgc gatctcagcc cagctgcaac ctctatctcc tgagctcaag 420
 cgatcctcct tagcttccca aatagctgga actacaggca tgtgccatca cgtccagcta 480
 attttgtatt tttagtagag aaggttttac catgttggac aggggtgtct cgaactcctg 540
 gcctaaagtg gtccacctag ctgagcctac caaagtgtg tgattacagg cgtgagccac 600
 catgcccagc ctctaaattc tgttttctat tcaaagtaaa aatgacatgt gtttgagtca 660
 aaaaaaaaaa aaaaaaagggt gggggtcggg aaagggcccg gggaattgtt cccggccatt 720
 caat 724

<210> 104
 <211> 734
 <212> DNA
 <213> Homo sapien

<400> 104
gagcggccgc cccggcaggc cccccccctt tttttttttt tttttttttt ttttttggtt 60
taaaaaaatt tgcctttgtt tttttctctt ttggcggggg ggccctgctt gaggggggat 120
ggggccccgg ggaatggggg ggctggggaa ataatccaca aagggtgtga aagagaaggg 180
gggtgtgttt tagaagcgc aggcgccagg gtgtcctcgt gtgccccgtt cttggggagg 240
ggacgcgcct ttgaggaagg gatttcttgg gctaacgcct atatgtctgg cgccccatt 300
ccatgtatta aaaatttctc tgagaaaatt tcttatgtg gccctattcg ttgttggtgg 360
ttgcccatgt tccttccaat acatgcgcgg tcaaggggac ccccgagggc cttctgcgg 420
tcccttggtg aagaagggg gaagatatgt ctctgttta atcactagta taaagccggt 480
ggcgtgcata tcaactcaagt gtgccatata tgccgggtct tctgggggtg tgatatatgt 540
gtgggcacct cccgcgctc caaacacacc ctttactaa caattccgtc gcgtgcacca 600
acagggcggt tttatcggag ggcagacgga gataagcggg ggataaaagg agatcaaaac 660
aagaaaggaa ggaaccgcaa aaaaacaaaa aaacaaacaa caaaaaaaaa aaaataagaa 720
gagcgagatg gagc 734

<210> 105
<211> 648
<212> DNA
<213> Homo sapien

<400> 105
aggatccgcc cgggcaggc cgggcaggta cctactaggc agttgggttc agggaaatag 60
ggattagact atggcctatc aggcctctat atggtcataa tatttaaaat ataggagta 120
gaaaacaaca aagaataggg ataggactac ttaaaaacaa tagaaagagc atatataac 180
gtatatagta cccgtatgaa tagtagaata tatagtatat tatagatata tcataaatat 240
actagctagg taacaatagt agacgagtta aacaataggc agcatataat agtaatataa 300
taatataata aatattacag aaataacgca ttattataaa tatattacta atataccagg 360
gtagacataa atagcattta aatattaggg atattagggt aggagtaatt aatagtatta 420
actaggagta tagtacaacg taaaatgaag gtcccccatc agcggaaaaa aaacaaaaac 480
acaaaaaaaa gaaaaaaaaa aaaaaagggt ttgtgggggg gttatactac gtgtgggcat 540
aatatagggtg tgttaccggg tgtgtgttgt gctagaacaa catgggtgtg tgtaataatt 600
acgggggggt tctcagacaa attttttcgc gacaaaaaaa atttatag 648

<210> 106

<211> 580
 <212> DNA
 <213> Homo sapien

<400> 106
 ttagactcat atagcgaatg tgtccattaa tcatgctcga gcggcgcgtg tgtgatggat 60
 gtcgcggcga ggtactagaa gtgaagcacc tttatttagc aataattaca aagagttgct 120
 taagattgat gcagataaat cattcatgaa actagaacaa gattatgaac tacattagta 180
 agttcccttca ttcagcaatt tatgccaaag atacactttc cctgacttca cttttccctg 240
 ccttgagata aaatgaggat aacagtggct atttcttagg gttgctataa agattaaatg 300
 agctgatact tgtaaagtat gtaaaagaag gcctgacata ttatcagttt ccattgacat 360
 ttctactttc aaggaacttg taatatagtt agggaggtaa catatgcaca taaacatcta 420
 aataaagatt ctcatgaaat gcccaagtaa gcaattctgt aatgtataaa aaaaaaaaaa 480
 aaaaaaaaaa aaaaaaaatt tttgtttttc tctttggtct ctttcttttt ccctttttta 540
 atttttttcc ccctcccaat ttcccccaa aatttgacaa 580

<210> 107
 <211> 1634
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (369)..(369)
 <223> a, c, g or t

<400> 107
 tcgcggcccg aggtaccctg tggcccagga ggacgtccgt acttccagcc ccgacacct 60
 ggacatctac tctgcccgtg tgatgactgc ctccattga tcatggaccc tgctgtcgag 120
 gtcgctggca atgaagttgc ttgaatacaa aatgactcca aggtacgtta atacgttgc 180
 taccacttc taattgagac agaataggta atggcccaac ggtccccaat gaagctgac 240
 gcccgcccat gacgcatgtg ttgcaggatg taggggtggg ggcgatagtg tatcggggg 300
 aagtctgtcg tgatcatgct cctccctgaa ggcatatccc ttgcgtgca gaatctccgt 360
 gtcgctccnt tgtgacgtgc ctctctgaca ggcgtcctct catttcacat cgtccctatg 420
 gacagtaccc acccaactac ccgtcataa ccttgtttcc tcgctccatg caagtgaatc 480
 ccaaccgaac cctattgcca tacacgcttg gagcacgcta ttataggggc ttgttgaatg 540
 acgataccta ggaaggtaaa agacgttctt ctataatatc tatccacttg cgtgcatatg 600
 ggggcaaaga gagagagcct acttatttct acacccttga ccttgggtct attgccaaat 660

ttaccaaact caaagaaacg tatacccaac gaatattatg taattggggg ttaccaaata 720
 catatacacc aaggaccata tattatataa acccaaagaa aaaactgagc ccggaagtgc 780
 aacgtcctta tcttaccctg gtgccaagtc ctattoccag gtcgtaatat caccctgttt 840
 ttgtgtgtgt agcccaaaact gtggcttagt actaccccc gggtgtccac gcgacaaaat 900
 tccccacac attaaacaaa gagaagtgtc gtcctatat tattacacac acaccaggg 960
 ggcccggccc ttatataaat tttttggggg cgccccttgg ggaaatttgg ggttctccac 1020
 acggggaact taaaatttcc aacccttta gaaaacgcca acgggttggg cacaggcccc 1080
 caaggcgtag actttacccc caacttttcg ctgtgtacgg gctggtctcc aactttatct 1140
 tttttggcct agggttcaca ccccaagga ccaaaaccgg gtttgagccc caacctgaaa 1200
 cggaaaaacc atttcctttt tggaacacag ggaaccaaca atttccttta gaaaccaatt 1260
 ttggaaaaag gccccaaac actgtggttt cccccctagt gggcaaacag cagcctttt 1320
 ccacttcca aaaaaggcct ttggagagcc ggtaaaact ccaatagggg ttcccaacta 1380
 aaggcttggc tttaccctt gggtgttttg acacactttg gtgtaatccg ccggctccca 1440
 caaattccca cacactcacc tcagatgaca tgcgagagca cagctgctcg cgcaaggagc 1500
 gagaaggcca actacacagc cgtaacactc atcagacgcg ccaggcgaca cagagcaac 1560
 aacactgctg gaacgagcac cacaatcagc gacgacgaga ctgaacgcag cgaccaacac 1620
 gctacgcagc agag 1634

<210> 108
 <211> 697
 <212> DNA
 <213> Homo sapien

<400> 108
 ctgatgcggc ctgcccgggc aggtcccccc cctttttttt tttttttttt tttttttttt 60
 tttgtttaaa aaaattgaac tttcgttttt ttccttttgg gcggtggtgg gcgcccatt 120
 aggggtggtg ggtgtcccca gggaaggggt gtggcgctgg gagaacatat gatctagcca 180
 gaaaagttgc ttgagaatga gaacgtggtg gtgtcgtgtt ttagaagtgc gccatgtggc 240
 caagggtggc gctcccctca ggctccgctt tctttggaga agtgtgagcc ccgcctgtag 300
 agagaaggag atctcattag cgcaaacgca caatacgcgt atgcgcacac acaatctcaa 360
 agattataaa agaaaatctc ttcgtagaaa caatctccta cgcgcgcccg ccactctca 420
 cgtcttgtgg cgtgtgtctc ccacgtattc tcatcatcat accatgtgcg cggtcacaag 480
 gtgtacaccc cgaggggttc tcccttctcg tgggtcttcc cgtgtgtgtg tgaagagggc 540

cactcataga ttccggtgta tactctacag tgaagacctg tgggtgttta tacactcggc 600
 ggtctcaata ccctttgtcc ccgtgggtgt gaaaatttgg ttaccgccc tcacaattct 660
 ccccaact tgcggaacaa aagatacacc gctgttt 697

<210> 109
 <211> 581
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (487)..(487)
 <223> a, c, g or t

<400> 109
 gcgagtgtgg cctctaagtc atgctcgagc ggcgcagtgt gatggattcg cggcgaggtc 60
 cccccctttt tttttttttt tttttttttt ttttttaggg agggggaaaa attttttttt 120
 tttttttatt ttctcccaa aacccttttt ttggaaaaaa ttaaaagttg caatgagggg 180
 ttttctttat aaaaaaata ttaaaactag gggcatacta ttattccaaa aaaagttaa 240
 tttgctatgt gttgacaaag cacatcacga gtgggtgtat aagctggtcc tctcttatat 300
 ttttcagaga aaatattatt ctacagtggt ccatgtctat tccatcacg tgtgttcaag 360
 gagaaaaatc cgtcgggcgt gtaactcact tggggtgcat aagtgtgtgt tacctctgtg 420
 tgaaatattg tgttttatcc ccgtccaca atattccac aacaacatct aaccagaaca 480
 cacacngtc tgcagcaaga ggcgggcgcg cagaggacaa gagacgggac aacgagcaag 540
 agaccaagca gcggaggcaa gagaaggaga gaggggaagta c 581

<210> 110
 <211> 862
 <212> DNA
 <213> Homo sapien

<400> 110
 ccgcccgggc aggtccctcc tccttttttt tttttttttt tttttttttt tttttttttt 60
 tttttttttt tttttttttt tttttgtggg ggtttttttt gtgggaaaaat tggagggggg 120
 ttaaaaaaat tcccccccc ccccttttaa aaccaccaata gtggggggcc cctggggggg 180
 gaaaaacagc ggtgggtggg tcgctgaaaa tgtagtctac taagtagata aaacagctgt 240
 gttcttgtgt ggtggcccca ccggtgttc cacatcttct attaatagat agtgtgtgtg 300
 ggtgccgcag ggaggcgcaa acaacatata ttttctttat ttcaaattca tttgtggggg 360

76

ggaaaaaac tttattgttc accacacatg cgtgggtaga tcacaacagc aaaagaagat 420
 gtgtcaaaat aaatgggtgt gctaaagaag ccgggtggcg tggaagacaa acactattag 480
 tgatcatgtt gtgtcggagt gtgtgtgatt atcctcccg cggggtaaga agagaggtgg 540
 tggctctgcaa cacaagagg ggcggcgga ggaggagaga acaaccatct atcaccgcg 600
 ttgcgcgct tatttacata tatatggtcg agggcgagat caaacatctc tcgagggaga 660
 gagagggcga gcgggcgaac ccaaccacgg cgggacaaca agaggccatc tccgcgggg 720
 aggaagaaag gggttgctcc gcgcggcgcg cccaacccc tccacacaac accctatacc 780
 gcacacaaca aaccaacca caacctcgga cacagtcaac aagagaatat aaaaaatat 840
 aacaaaaaaa acaacaaaag tg 862

<210> 111
 <211> 298
 <212> DNA
 <213> Homo sapien

<400> 111
 cacaacatac gagcattgct gaagaaaaa aatcatcaga agctcttcag gaggtgtgtc 60
 gatatggaca gacacagaat ccagaatcc caccacaata agatggagca attccaaaat 120
 aagggctaata ggagcccgaa aggtatcatc gccagcatgg caacaagaat gagccaacag 180
 gccgacaaag attgaacgga taatcatagc caccatacga agttctcatg actgtacggg 240
 aatacataga caaatcaaac atacgctaca actgtccaag ggaaacatca ttccagga 298

<210> 112
 <211> 638
 <212> DNA
 <213> Homo sapien

<400> 112
 tgatgatcga tatataggcc aatggttcat ctaatcatgc tcgagcggcg cagtgtgatg 60
 gatcgtgtgc gggccgagg tccccccctt tttttttttt tttttttttt tttttttttt 120
 tttttttaaa atttgtaaat tttttttttg tctcttcccc ccaaaacccc ctttttttaa 180
 atttttacgt ttttcggggt ttttttaaaa aaaaattaaa acagctttaa acaccttttt 240
 ttaaaacgat tatttcaaag aactattttg tcgaaagaaa tactccggga tcccgatatt 300
 atcgagcgtc ccatctttat atttatcaaa ataataattct agccagggac cgtgcgtttc 360
 taagccgatc cggggtgtgc gtcggagcat aaactatggg gaggcaatag caggcagtga 420
 ggcagctcgg aggtctgtat acatcggtcg ttgtggagta acatgtgtca ttgtccgcgc 480
 gtcccaccaa ttccgaagc aacaaacttt gtaacgagag agcagataca ggagatcagc 540

tcgcaagcga aagtccagtc agcggctaac cacggcagac acgccgagcc aagacgtcgt 600
tgtgcatcgc tagtgcctgg tagcgacggg gcggcgct 638

<210> 113
<211> 783
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (304)..(304)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (403)..(403)
<223> a, c, g or t

<220>
<221> misc_feature
<222> (622)..(622)
<223> a, c, g or t

<400> 113
gatgatacga ctactatagg cgaattggtc atctagatca tgctcgaagc ggcgcagtgt 60
gatggatcgg gagcgggctg cggggcaggt cccccccct tttttttttt tttttttttt 120
ttttttccct gtaaagattt tttttttttc ctctaaaaaa gtccactttt aaaatggggt 180
tcccggaaaa tttaccaggt ggctcttttt aaaaggggca aaagggttgc attccaattc 240
cgggggtttg tttcccccat cccaatttt tgggggctgt ggcaaaaacg gcggctctta 300
gggnaaagag gaggggttgt ttaaaggag acagaggagt ggtataaaac acccggttgc 360
ttgttgtag gaactcatca atataccata tttctctac agntgagtgt ggcttcatta 420
ttggggctctg ttcggccgtg gcacccccct tcagatagtg tgtgatttga gagagacaag 480
caaggagggg ccgacgtgtc tccattatct ctacagacac caccggtggg ggtgcgggtg 540
cgcgctgcct tgccaaaggg agaaaagggg gggcttatgg cgcgcacacc tttctcaaca 600
agaaacaccg agcccccca antgattggg gccagtaatg atgaggccct ggggtgggata 660
ctcatggtgg cacataaggc gtcgtctccg ggtgttgacc agtgtgttac tcccgtcac 720
aatcccccca aaacatggca ccaacaaaaa catgagagga cgacagcaca cagaacgaag 780
aag 783

<210> 114
<211> 648
<212> DNA
<213> Homo sapien

<400> 114
acaagcttat aggactctat gacaggctgt gaatgttttt ttgttggttg ttgttgtttt 60
taattgctgt taatattttt taaataataa agaaacaaaa ctagaataaa aaaaaaaaaa 120
aaaaaaaaaa aaaggtgtgg gacttggggg atgtggtgga agggaatata cggtgcccca 180
ttatctttta aaccgtgtgt tccccttttt aaataccggg gattattttt ttccaaggga 240
cagttttttt aaagaaaact ttggagagtg ggggaggaac cacatggggc aaaacggcgt 300
gtccccgggt gggaaatgtg ggtgcaccgg gctcaaaatt cccaccaaac aattcgagac 360
aacgaaaaac gaacagcaac aggagaaaga agaacaaca cgacacacac gaaacagaag 420
gagaagagaa agagagagaa acaccaacac acagcaacca agaaaagacg aaaaagaaag 480
ggaaaaaaga gaaagaaaag aagaaaaaag agaaaacaag aagaaagaac accagaaaga 540
aaagaaaaac acaaagacaa gacaacacac aaaacaaaga aaaacagggc gaacaacaaa 600
agaagacaaa aacagcaacg aaaaacagga gagaactaaa acaaagag 648

<210> 115
<211> 928
<212> DNA
<213> Homo sapien

<400> 115
aagaattcgg cacgagggtc cctccctagg tcctcaaggc cctccaggct atggcaagat 60
gggtgcaaca ggaccaatgg gccagcaagg catccctggc atccctgggc ccccggtcc 120
catgggccag ccaggcaagg ctggccactg taatccctct gactgctttg gggccatgcc 180
gatggagcag cagtaccac ccatgaaaac catgaagggg ccttttggct gaaattcccc 240
acctgccttt ggatgaaaga ctccgttggg aataaatggc caaagcttat aggactctgt 300
gacaggttgt gaatgttttt ttgttggttg ttgttgtttt taattgctgt taatattttt 360
taaataataa agaaacaaaa ctaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaggtgtgg 420
gacttggggg atgtggtgga agggaatata cggtgcccca ttatctttta aaccgtgtgt 480
tccccttttt aaataccggg gattattttt ttccaaggga cagttttttt aaagaaaact 540
ttggagagtg ggggaggaac cacatggggc aaaacggcgt gtccccgggt gggaaatgtg 600
ggtgcaccgg gctcaaaatt cccaccaaac aattcgagac aacgaaaaac gaacagcaac 660
aggagaaaga agaacaaca cgacacacac gaaacagaag gagaagagaa agagagagaa 720

79

acaccaacac acagcaacca agaaaagacg aaaaagaaag ggaaaaaaga gaaagaaaag 780
 aagaaaaaag agaaaaacaag aagaagaac accagaaaga aaagaaaaac acaaagacaa 840
 gacaacacac aaaacaaaga aaaacagggc gaacaacaaa agaagacaaa aacagcaacg 900
 aaaaacagga gagaactaaa acaaagag 928

<210> 116
 <211> 82
 <212> PRT
 <213> Homo sapien

<400> 116

Met Met Arg Glu Ser Phe Phe Val Leu Ala Val Leu Ile Ile Leu Gly
 1 5 10 15

Gly Ala Thr His Pro Pro Pro Pro Pro Cys Ser Thr Pro Ala Val
 20 25 30

Val Phe Pro Pro Ser Leu Val Gln Pro Val Phe Ile Met Thr Cys Cys
 35 40 45

Tyr His Val Val Leu Leu Phe Val Ala Pro Leu Cys Gly Gly Pro Pro
 50 55 60

Pro Leu Glu Arg Ala Ser Pro Val Pro Phe Val Gly Arg Gln Gln Gln
 65 70 75 80

Ser Ala

<210> 117
 <211> 35
 <212> PRT
 <213> Homo sapien

<400> 117

Met Val Phe Phe Phe Phe Phe Phe Lys Lys Trp Ser Leu Cys Asn
 1 5 10 15

Phe Ala Lys Val Asp Phe Glu Phe Arg Gly Pro Ile Asp Pro Pro Thr
 20 25 30

Ser Ala Ser
 35

80

<210> 118
 <211> 107
 <212> PRT
 <213> Homo sapien

<400> 118

Met Tyr Leu Gly Pro Leu Arg Asn Leu Leu Asp Val Ser Lys Lys Lys
 1 5 10 15

Lys Lys His Pro Gln Lys Glu Gln Pro Arg Gly Ala Leu Glu Cys Gly
 20 25 30

Ser Pro Leu Ser Val Val Leu Cys Phe Ser Pro Ile Ser Phe Leu Glu
 35 40 45

Ala Arg Glu Gly His Pro Ser Val Gly Ser Ser Thr Ile Leu Leu Glu
 50 55 60

Ala Ser His Ser Pro Ala Phe Leu Leu Leu Pro Lys Pro Val Phe Leu
 65 70 75 80

Leu His Leu Gly Glu Gly Gly Lys Cys Val Pro Gly Leu Glu Asn Trp
 85 90 95

Cys Leu Thr Gly Lys Val Ser Gly Pro Pro Arg
 100 105

<210> 119
 <211> 75
 <212> PRT
 <213> Homo sapien

<400> 119

Met Ala Thr Pro Val Phe Gln Leu Leu Arg Pro Arg Thr Leu Gly Tyr
 1 5 10 15

Leu Arg Thr Leu Leu Leu Ser Phe Pro Met Ser Gly Glu Ser Leu Ser
 20 25 30

Phe Val Asp Cys Ala Thr Lys Met Tyr Leu Glu Ser Asp His Ile Ser
 35 40 45

Gly Thr Ser Ala Ala Thr Arg Ile His His Asn Leu Ala Ala Ala Glu
 50 55 60

Gln His Leu Gly Asp Thr Ser Pro His Arg His

81

65

70

75

<210> 120
 <211> 195
 <212> PRT
 <213> Homo sapien
 <400> 120

Met Ala Pro Gly Tyr Pro Pro Ser Phe Leu Lys Lys Lys Trp Leu Leu
 1 5 10 15

Glu Asn Lys Arg Arg His Pro Arg Lys Leu Gly Glu Glu Thr Thr Phe
 20 25 30

Cys Pro Ser Pro Pro Tyr Gly Gly Leu Arg Glu Pro Thr Gly His Arg
 35 40 45

Gln Pro Leu Phe Ser Leu Asp Arg Ala His Glu Lys Val Pro Pro Arg
 50 55 60

Arg Tyr Ile Val Leu Val Gly Thr Gln Ala Ser Gly Pro Val Val Arg
 65 70 75 80

Val Arg Asp Asn Thr Leu Gly Lys Lys Asn Lys Ser Asn Asn Leu Val
 85 90 95

Leu Leu Leu Ala Tyr Arg Thr Arg Lys Arg Asn Thr Arg Ser Arg Leu
 100 105 110

Arg Leu Ser Gln His Met Arg Glu Lys Ala Leu Gln Thr Trp Leu Glu
 115 120 125

Ser Trp Thr Phe Val Lys Gly Glu Lys Ile Val Pro Ala Pro His Val
 130 135 140

Leu Leu Thr Ala Leu Arg Ser Thr Gly Asn Pro Gln Arg Lys Gly Gly
 145 150 155 160

Gly Glu Ser Trp Val Leu Gly Trp Glu Gln Leu Cys Gly Thr Pro Pro
 165 170 175

Glu Leu Arg Val Trp Val Lys Gly Ser His Asn Ser Phe Phe Lys Lys
 180 185 190

Asn Lys Phe

82

195

<210> 121
 <211> 36
 <212> PRT
 <213> Homo sapien

<400> 121

Met Ser Cys Phe Phe Phe Ala Phe Leu Lys Met Glu Val Thr Ala Lys
 1 5 10 15

Trp Glu Ile Asn Leu Pro Ile Asn Ser Cys Asn Met Thr Thr Ala Glu
 20 25 30

Gln Cys Leu Glu
 35

<210> 122
 <211> 117
 <212> PRT
 <213> Homo sapien

<400> 122

Met Leu Arg Gly Ala Arg Glu Thr His Ile Ser Thr His His Ala Trp
 1 5 10 15

Asn Thr Ala Leu Leu Glu Thr Thr Arg Asp Val Tyr Pro Pro Gln Leu
 20 25 30

Ser Cys Leu Gly Gly Glu Arg Lys Ile Trp Leu Val Arg Gln Gly Gly
 35 40 45

Phe Val Pro His Leu Arg Gly Gly Gly Glu Asn Ile Pro Arg Leu Val
 50 55 60

Phe Val Tyr Lys Thr Asn Lys Cys Lys Lys Leu Ser Thr Asn Phe Phe
 65 70 75 80

Gly Thr Lys Gly Val Gly Val Ser Arg Arg Ser Phe Ala His Gly Thr
 85 90 95

Ala Glu Trp Ser Gln Ser Ser Val Glu Thr Lys Ile His Phe Ala Ser
 100 105 110

Thr Phe Lys Pro Val
 115

83

<210> 123
 <211> 10
 <212> PRT
 <213> Homo sapien

<400> 123

Met Gly Arg Ser Leu Glu Val His Gly Val
 1 5 10

<210> 124
 <211> 42
 <212> PRT
 <213> Homo sapien

<400> 124

Met Trp Arg Lys Gln Phe Pro Pro Gly Glu Thr Val Trp Pro Gly Phe
 1 5 10 15

Pro Pro Gly Phe Phe Phe Phe Leu Leu Cys Phe Phe Gly Asn Ser Phe
 20 25 30

Met Thr Phe Asn Leu Thr Met Asn Tyr Gln
 35 40

<210> 125
 <211> 315
 <212> PRT
 <213> Homo sapien

<400> 125

Phe Tyr Tyr Lys Thr Lys Ile Thr Lys Thr Gly Trp Tyr Trp His Lys
 1 5 10 15

Asp Lys His Leu Asp Gln Ala Asn Arg Ile Glu Thr Ala Glu Val Asn
 20 25 30

Ser Tyr Ile Tyr Leu Gln Leu Asn Phe Tyr Lys Gly Val Arg Thr Ile
 35 40 45

Pro Ser Glu Asn Asn Ile Phe Asn Lys Ser Leu Trp Val Asn Cys Ile
 50 55 60

Asp Thr Cys Lys Thr Met Lys Leu Asp Ser Ala His Ile Leu Tyr Ala
 65 70 75 80

Lys Ile Asn Phe Asn Ala Leu Gln Thr Ala Ile Gln Glu Leu Lys Leu
85 90 95

Lys Ile Ile Glu Glu Lys Val Arg Val Thr Leu His Asp Leu Ala Phe
100 105 110

Asn Asn Glu Leu Ser Ile Met Ile Pro Lys Thr Gln Ala Ile Lys Asn
115 120 125

Lys Lys Asp Lys Arg Gln Pro Thr Lys Trp Glu Lys Ile Cys Ala Asn
130 135 140

Tyr Ile Ser Asn Lys Asp Leu Leu Ser Arg Leu Ala Leu Leu Gln Pro
145 150 155 160

Tyr Thr Lys Thr Ala Leu Ile Ala Lys Leu Pro Lys Asp Leu Asn Arg
165 170 175

His Phe Phe Lys Glu Asp Ile Leu Val Ala Gln Lys His Met Lys Arg
180 185 190

Cys Ser Ile Ser Leu Ile Ile Arg Glu Met Gln Ile Lys Ser Pro Met
195 200 205

Arg Tyr His Phe Thr Pro Thr Arg Met Ala Ile Ile Lys Lys Lys Thr
210 215 220

Glu Asn Asn Lys Gly Phe Ser Gly Cys Gly Glu Ile Cys Asn Phe Ile
225 230 235 240

His Cys Trp Ala Glu Tyr Thr Met Ala Gln Pro Pro Trp Arg Thr Val
245 250 255

Trp Glu Val Leu Gln Lys Val Glu Gln Asn Tyr Asn Met Thr Gln Gln
260 265 270

Ile Pro Leu Leu Asp Ile Tyr Pro Gln Lys Asn Lys Thr Cys Cys Pro
275 280 285

Leu Lys Pro Cys Thr Gln Met Phe Thr Ala Ile Leu Phe Ile Ile Ala
290 295 300

Lys Lys Lys Val Glu Thr Thr Asn Gln Trp Ile
305 310 315

85

<210> 126
 <211> 66
 <212> PRT
 <213> Homo sapien

<400> 126

Met Phe Leu Pro Tyr Gly Lys Ser Glu Ala Ala Arg Glu Ala Ser Gly
 1 5 10 15

Ala Cys Lys Thr Thr Asp Gly Ile Val Ser Glu Leu Thr Met Asn Thr
 20 25 30

Cys Ser Pro Leu Ser Ile Asp Gln Ser Lys Ser Asn Val Val Gly Lys
 35 40 45

Gly Pro Ser Pro Thr Val Gly Gly Glu Gly Cys Gly His Leu Pro Leu
 50 55 60

Ala Asp
 65

<210> 127
 <211> 40
 <212> PRT
 <213> Homo sapien

<400> 127

Met Glu Thr Lys Tyr Val His His Gln His Ile Phe Tyr Tyr Arg Leu
 1 5 10 15

Pro Asn Ile Arg Phe Thr Asn Phe Ser Asn Phe Pro Thr Arg Asp Leu
 20 25 30

Ser Phe Asn Val Pro Arg Asn Tyr
 35 40

<210> 128
 <211> 80
 <212> PRT
 <213> Homo sapien

<400> 128

Met Gly Val Gly Ala Gly Arg Thr Phe Phe Thr Arg Gly Pro Ser Ser
 1 5 10 15

Gly Pro Val Val Arg Arg Asn Ala Leu Pro Phe Phe Phe Leu Lys Lys

86

20

25

30

Gly Val Ser Cys Leu Phe Cys His Arg Leu Gly Gly His Asn Trp Glu
 35 40 45

Gln Ile Val Gly Gly Ser Val Ile Arg Phe His Pro Pro Thr Gly Val
 50 55 60

Tyr Ser Ala Ile Leu Pro Val Ala Arg Leu Pro Cys Leu Pro Trp Arg
 65 70 75 80

<210> 129
 <211> 88
 <212> PRT
 <213> Homo sapien

<400> 129

Met Tyr Leu Ser Phe Met Ser Pro Arg Arg His Thr Gln Lys Val Lys
 1 5 10 15

Ser Pro Gly Arg Gly Leu Arg Ser Leu Pro Ser Gly Leu Pro Pro Ala
 20 25 30

Arg Glu Ala Pro Gln Cys Gly Arg Pro Leu Pro Arg Pro Thr Pro Arg
 35 40 45

Leu Cys Pro Val Pro Thr Leu Ala Val Trp Ala Thr Pro Ser Glu Leu
 50 55 60

Leu Glu Ala Thr Asn Thr Gln Val Ser Tyr Ser Thr Ser Thr Asp Pro
 65 70 75 80

Gly Leu Met Gly Leu Tyr Ile Lys
 85

<210> 130
 <211> 49
 <212> PRT
 <213> Homo sapien

<400> 130

Met Asn Gln Asn Arg Gly Ser Gln Ser Arg Glu Lys Lys Ile Leu Gly
 1 5 10 15

Ser Glu Ser Thr Leu Cys Pro Phe Glu Leu Gln Lys Glu Lys Glu Thr
 20 25 30

87

Lys Ala Lys Ser Asn Gly Gly Gln Ala Ala Arg Tyr Leu Pro Gly Arg
 35 40 45

Arg

<210> 131
 <211> 87
 <212> PRT
 <213> Homo sapien

<400> 131

Met Val Val Phe Val Ser Cys Met Tyr Arg Phe Cys Ser Leu Arg Leu
 1 5 10 15

Leu Thr Val Gly Arg Arg His Lys Met Gly Ala Asp Cys Phe Ser His
 20 25 30

Asn Ile Cys Gly Gly Asn Cys Gly Ala Gly Met Thr Pro His Phe Gln
 35 40 45

His Gln Gly Thr Ser Val Met Ala His Glu Phe Ser Val Pro Ser Phe
 50 55 60

Ser Cys Glu Ser Gln Asp Ile Ser Cys Ala Phe Ser His Lys Asp Thr
 65 70 75 80

Arg Glu Gly Pro Gly Val His
 85

<210> 132
 <211> 26
 <212> PRT
 <213> Homo sapien

<400> 132

Met Leu Ser Ser Gly Ala Val Val Met Ile Glu Arg Arg Pro Gly Gln
 1 5 10 15

Val Leu Ala Leu Lys Thr Ile Thr Lys Gln
 20 25

<210> 133
 <211> 519
 <212> PRT

88

<213> Homo sapien

<400> 133

Met Thr Cys Pro Asp Lys Pro Gly Gln Leu Ile Asn Trp Phe Ile Cys
 1 5 10 15

Ser Leu Cys Val Pro Arg Val Arg Lys Leu Trp Ser Ser Arg Arg Pro
 20 25 30

Arg Thr Arg Arg Asn Leu Leu Leu Gly Thr Ala Cys Ala Ile Tyr Leu
 35 40 45

Gly Phe Leu Val Ser Gln Val Gly Arg Ala Ser Leu Gln His Gly Gln
 50 55 60

Ala Ala Glu Lys Gly Pro His Arg Ser Arg Asp Thr Ala Glu Pro Ser
 65 70 75 80

Phe Pro Glu Ile Pro Leu Asp Gly Thr Leu Ala Pro Pro Glu Ser Gln
 85 90 95

Gly Asn Gly Ser Thr Leu Gln Pro Asn Val Val Tyr Ile Thr Leu Arg
 100 105 110

Ser Lys Arg Ser Lys Pro Ala Asn Ile Arg Gly Thr Val Lys Pro Lys
 115 120 125

Arg Arg Lys Lys His Ala Val Ala Ser Ala Ala Pro Gly Gln Glu Ala
 130 135 140

Leu Val Gly Pro Ser Leu Gln Pro Gln Glu Ala Ala Arg Glu Ala Asp
 145 150 155 160

Ala Val Ala Pro Gly Tyr Ala Gln Gly Ala Asn Leu Val Lys Ile Gly
 165 170 175

Glu Arg Pro Trp Arg Leu Val Arg Gly Pro Gly Val Arg Ala Gly Gly
 180 185 190

Pro Asp Phe Leu Gln Pro Ser Ser Arg Glu Ser Asn Ile Arg Ile Tyr
 195 200 205

Ser Glu Ser Ala Pro Ser Trp Leu Ser Lys Asp Asp Ile Arg Arg Met
 210 215 220

Arg Leu Leu Ala Asp Ser Ala Val Ala Gly Leu Arg Pro Val Ser Ser
 225 230 235 240

Arg Ser Gly Ala Arg Leu Leu Val Leu Glu Gly Gly Ala Pro Gly Ala
 245 250 255

Val Leu Arg Cys Gly Pro Ser Pro Cys Gly Leu Leu Lys Gln Pro Leu
 260 265 270

Asp Met Ser Glu Val Phe Ala Phe His Leu Asp Arg Ile Leu Gly Leu
 275 280 285

Asn Arg Thr Leu Pro Ser Val Ser Arg Lys Ala Glu Phe Ile Gln Asp
 290 295 300

Gly Arg Pro Cys Pro Ile Ile Leu Trp Asp Ala Ser Leu Ser Ser Ala
 305 310 315 320

Ser Asn Asp Thr His Ser Ser Val Lys Leu Thr Trp Gly Thr Tyr Gln
 325 330 335

Gln Leu Leu Lys Gln Lys Cys Trp Gln Asn Gly Arg Val Pro Lys Pro
 340 345 350

Glu Ser Gly Cys Thr Glu Ile His His His Glu Trp Ser Lys Met Ala
 355 360 365

Leu Phe Asp Phe Leu Leu Gln Ile Tyr Asn Arg Leu Asp Thr Asn Cys
 370 375 380

Cys Gly Phe Arg Pro Arg Lys Glu Asp Ala Cys Val Gln Asn Gly Leu
 385 390 395 400

Arg Pro Lys Cys Asp Asp Gln Gly Ser Ala Ala Leu Ala His Ile Ile
 405 410 415

Gln Arg Lys His Asp Pro Arg His Leu Val Phe Ile Asp Asn Lys Gly
 420 425 430

Phe Phe Asp Arg Ser Glu Asp Asn Leu Asn Phe Lys Leu Leu Glu Gly
 435 440 445

Ile Lys Glu Phe Pro Ala Ser Ala Val Ser Val Leu Lys Ser Gln His
 450 455 460

90

Leu Arg Gln Lys Leu Leu Gln Ser Leu Phe Leu Asp Lys Val Tyr Trp
 465 470 475 480

Glu Ser Gln Gly Gly Arg Gln Gly Ile Asp Lys Leu Ile Asp Val Ile
 485 490 495

Glu His Arg Ala Lys Ile Leu Ile Thr Tyr Ile Asn Ala His Gly Val
 500 505 510

Lys Val Leu Pro Met Asn Glu
 515

<210> 134
 <211> 66
 <212> PRT
 <213> Homo sapien

<400> 134

Met Gly Arg Asp Lys Ser Glu Val Thr Val Asn Asn Lys Val Met Phe
 1 5 10 15

Tyr Gly Tyr Phe Ile Gly Asp Lys Phe Ile Thr Arg Ala Ile Ser Tyr
 20 25 30

His Val Leu Ile Leu Pro Gly Cys Asn Met Val Thr Leu Glu Thr Gln
 35 40 45

Ile Leu Asn Ile Gly Val Lys Thr Thr Ser Cys His Ser Ile Leu Ser
 50 55 60

Thr Val
 65

<210> 135
 <211> 91
 <212> PRT
 <213> Homo sapien

<400> 135

Met Val Cys Val Val Val Ala Cys Gly Trp Ala Asp Val Cys Val Pro
 1 5 10 15

Ser Trp Cys Val Leu Cys Cys Ser Val Val Ser Trp Leu Val Val Val
 20 25 30

Arg Asp Thr Lys Thr Ser Thr Gln Arg Gly Gly Val Thr Arg Thr Asn
20 25 30

92

Ser Gly Ala Asp Glu Pro His Asn Lys His Ile Glu Thr Glu Ile Thr
 35 40 45

Lys Thr Asp Thr Asn Asn Arg Asp Thr Gln Arg Thr Lys Gln Ala Gln
 50 55 60

Lys Pro Asn Lys Glu Glu Ala Arg Lys Ala Gln Pro Thr Ser Thr Thr
 65 70 75 80

Thr Asn Lys Thr Asn Gly Thr Lys Glu His Ser Lys Gln Gln Thr Pro
 85 90 95

Thr His Asn His Thr
 100

<210> 138
 <211> 80
 <212> PRT
 <213> Homo sapien

<400> 138

Met Val Cys Ala Ala Trp Leu Pro Ser Ala Cys Pro Pro Trp Ser Val
 1 5 10 15

Asp Ala Pro Ser Thr Pro Leu Leu Gly Pro Cys Gln Pro Leu Val Val
 20 25 30

Glu Phe Ser Ser Pro Gly Val Val Val Gly Gly Pro Ser Met Ser Val
 35 40 45

Trp Gly Gly Arg Leu Arg Cys Pro His Trp Met Gln Pro Phe Ser Thr
 50 55 60

Ile Ser Gly Leu Lys Arg Asp Arg Val Arg Asn Val Asp Pro Leu Val
 65 70 75 80

<210> 139
 <211> 36
 <212> PRT
 <213> Homo sapien

<400> 139

Met His Leu Glu Arg Arg Ser Val Met Asp Gly Glu Val Asn Leu Ile
 1 5 10 15

Ser Leu Ser Gly Phe Leu Ser Tyr Cys Ile Phe Ile Tyr Lys Thr Asn

93

20

25

30

Phe Ile Leu Lys
35

<210> 140
<211> 45
<212> PRT
<213> Homo sapien

<400> 140

Met Trp Asn Phe Val Phe Leu Leu Ile Gly Gly Gly Gly Leu Ile Arg
1 5 10 15

Gly Val Val Cys Ala Pro Arg Arg Met Val Gly Val Cys Glu Asn Asn
20 25 30

Lys Lys Asn Val Leu Arg Arg Glu Arg Gly Val Val Cys
35 40 45

<210> 141
<211> 136
<212> PRT
<213> Homo sapien

<400> 141

Met Gly Trp Asn Thr Val Gly Arg Ser Gln Leu Ser Ala Ala Leu Asn
1 5 10 15

Ser Trp Ala Gln Ala Met Phe Ser Pro Gln Leu Pro Ser Ser Trp Ala
20 25 30

Cys Arg His Val Ser Ala Cys Leu Ala Tyr Phe Leu Phe Phe Phe Phe
35 40 45

Ser Phe Phe Phe Phe Leu Phe Phe Phe Phe Tyr Phe Phe Phe Leu Leu
50 55 60

Lys Arg Ala Gly Gly Gly His Ile Met Val Trp Arg Arg Arg Arg Trp
65 70 75 80

Ser Leu Gln Thr Ser Gly Val Pro Glu Val Val Phe Ser Ala Glu Cys
85 90 95

Cys Val Thr Thr Arg Cys Arg Gly Ser Thr Arg Trp Gly Lys Glu Ser
100 105 110

Val Ala Trp Gly Arg Asn Val Val Val Ala Arg Pro Asn Phe Ala Pro
 115 120 125

Lys Ile Ala Arg Thr His Glu Asn
 130 135

<210> 142
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 142

Met Asp Gln Ile Phe Pro Lys Arg Tyr Leu Met His Asn Ala Lys Lys
 1 5 10 15

Thr Lys Lys Lys Lys Lys Arg Gly Gly Lys Pro Ala Gln Glu Arg Ala
 20 25 30

Arg Gly Glu Thr Gly Val Pro Gly Pro Asn Phe Pro Lys Lys Phe Ala
 35 40 45

Thr Gln Lys
 50

<210> 143
 <211> 219
 <212> PRT
 <213> Homo sapien

<400> 143

Met Val Leu Ala Leu Ile Val Asp Leu Cys Leu Trp Leu Ser Pro Arg
 1 5 10 15

Thr Gly Ala Gly Arg Leu Thr Ser Phe Leu Ser Leu Ser Leu Cys Arg
 20 25 30

Leu Ser Leu Cys Leu Phe Tyr Leu Phe Gly Val Ser Gly Gly Trp Cys
 35 40 45

Gly Asp Ser Ser Ser Phe Ser Val Leu Pro Pro Arg Ile Arg Phe Arg
 50 55 60

Gly Arg Arg Ala Ala Val Val Ala Ser His Leu Leu Ile Ser Ala Pro
 65 70 75 80

95

Pro Leu Phe Cys Val Val Phe Leu His Cys Cys Ser Ala Val Cys Ser
85 90 95

Ser Trp Arg Arg Val Ser Gly Leu Cys Arg Pro Pro Leu Leu Arg Ser
100 105 110

Ser Arg Phe Cys Arg Arg Pro Leu Leu Leu Ser Phe Ile Thr Pro His
115 120 125

Leu Ser Ser Ser Arg Arg Gly Val Val Thr Phe Gly Leu Val Leu Pro
130 135 140

Phe Phe Trp Trp Leu Gly Arg Arg Ala His Asp Phe Phe Val Ser Pro
145 150 155 160

Arg Trp Leu Gly Ala Pro Gly Pro Pro Lys Lys Lys Pro Pro Pro Pro
165 170 175

Pro Thr Pro Gln Lys Lys Lys Thr Pro Pro Pro Pro Pro Lys Lys Lys
180 185 190

Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys
195 200 205

Lys Lys Lys Gly Gly Gly Thr Ser Ala Ala Thr
210 215

<210> 144
<211> 37
<212> PRT
<213> Homo sapien

<400> 144

Met Arg Ser Phe Arg Glu Ile His Ser Glu Arg Thr Leu Met Val Asn
1 5 10 15

Leu Arg Gly Lys Ser Gln Asp Ala Gln Lys Leu Trp Ser Leu Val Leu
20 25 30

Ile Ser Gln Ser Ile
35

<210> 145
<211> 280
<212> PRT

<213> Homo sapien

<400> 145

```

Met Val Val Phe Gly Val Ile Cys Leu Cys Cys Val Cys Pro Ile Leu
1           5           10           15

Phe Phe Ser Val Phe Leu Phe Val Val Val Cys Ser Val Val Cys Leu
           20           25           30

Leu Ser Leu Val Ser Ala Gly Cys Leu Val Gly Glu Leu Pro Phe Cys
           35           40           45

Phe Ser Phe Val Leu Cys Val Leu Gly Arg Ala Leu Ser Leu Leu Pro
50           55           60

Ser Leu Val Val Trp Leu Leu Ser Ser Ser Leu Cys Val Ser Leu Trp
65           70           75           80

Ser Phe Leu Leu Phe Leu Val Leu Val Val Leu Val Ser Arg Gly Phe
           85           90           95

Phe Ser Phe Val Ser Gly Ile Cys Val Cys Val Leu Cys Leu Leu Ser
           100          105          110

Phe Val Phe Val Val Cys Cys Arg Leu Arg Leu Phe Ile Ser Arg Leu
           115          120          125

Cys Leu Leu Arg Phe Leu Tyr Leu Ser Ser Val Cys Phe Ser Leu Phe
           130          135          140

Phe Ser Phe Ala Val Val Ser Arg Val Leu Phe Pro Thr Arg Gly Cys
145           150          155          160

Val Leu Leu Trp Leu Arg Gly Asp Thr Gln Ile Leu Trp Gly Gly Lys
           165          170          175

Val Cys Gly Arg Arg Pro Arg Gly Asp Thr Pro His Met Met Phe Pro
           180          185          190

His Pro His Ala Gly Leu Ile Thr Ala Leu Phe Gly Ala Pro Thr Arg
           195          200          205

Gly Val Tyr Ser Pro Pro Thr Ala Arg Phe Phe Val Val Tyr Ile Ile
210           215          220

```

97

Gly Asp Thr Ser Phe Phe Arg Gly Gly Pro His His Tyr Leu Gly Gly
 225 230 235 240

Ser Thr His Leu Gly Glu Thr Pro Arg Ala Val Ser Ser Leu Ile Ile
 245 250 255

Tyr Ile Lys Ile Tyr Gly Ala Arg Asp Arg Arg Tyr Ile Thr Arg Gly
 260 265 270

Leu Ser Phe Val Asp Ser Glu Lys
 275 280

<210> 146
 <211> 95
 <212> PRT
 <213> Homo sapien

<400> 146

Met Pro Val Val Pro Ala Ile Trp Glu Ala Lys Glu Asp Arg Leu Ser
 1 5 10 15

Ser Gly Asp Arg Gly Cys Ser Gly Leu Arg Ser Ala Pro Gln Pro Ser
 20 25 30

Ser Leu Val Lys Arg Glu Arg Phe His Arg Leu Ile Asn Gln Gln Thr
 35 40 45

Lys Thr Arg Ile Tyr Asp Gln Ala Gln Trp Leu Thr Pro Ile Ile Pro
 50 55 60

Val Leu Trp Glu Ala Arg Ala Gly Arg Phe Phe Glu Val Arg Ser Ser
 65 70 75 80

Arg Pro Ala Trp Ala Thr Gln Gly Asp Pro Val Ser Thr Lys Val
 85 90 95

<210> 147
 <211> 90
 <212> PRT
 <213> Homo sapien

<400> 147

Arg Ile Tyr Asp Gln Ala Gln Trp Leu Thr Pro Ile Ile Pro Val Leu
 1 5 10 15

98

Trp Glu Ala Arg Ala Gly Arg Phe Phe Glu Val Arg Ser Ser Arg Pro
 20 25 30

Ala Trp Ala Thr Gln Gly Asp Pro Val Ser Thr Lys Ser Leu Lys Ile
 35 40 45

Ser Ala Val Trp Trp His Thr Ser Val Val Ser Pro Thr Leu Glu Ala
 50 55 60

Glu Val Asp Cys Ser Ser Pro Gly Val Gln Ala Ser Val Ser Tyr Asp
 65 70 75 80

His Ser Thr Ala Leu Pro Ala Arg Gln Glu
 85 90

<210> 148
 <211> 79
 <212> PRT
 <213> Homo sapien

<400> 148

Met Ser Ser Leu Leu Pro Ala Phe Phe Val Ser Ile Asn Val Thr Ser
 1 5 10 15

Thr Tyr Pro Val Ile Gln Gly Lys Thr Gln Trp Arg Lys Pro Ser Ser
 20 25 30

Thr Thr His Ser Leu Tyr Leu Thr Leu Ser Gln His Pro Ala Lys Ser
 35 40 45

Arg Ser Lys Tyr Ser Ser Ser Leu Ser Thr Ser Leu Pro Phe Leu Gln
 50 55 60

Ser Ile Thr Leu Val Tyr Ser Ile Thr Ile Ser Gln Leu Asp Tyr
 65 70 75

<210> 149
 <211> 32
 <212> PRT
 <213> Homo sapien

<400> 149

Met Gly Ser Thr Thr Asp Val Ser Gly Ser Gln Cys Gly Cys Gln Phe
 1 5 10 15

Leu Tyr Leu Ala Ala Thr Thr Leu Ser Ile Thr Leu Arg Arg Ser Arg

99

20

25

30

<210> 150
<211> 57
<212> PRT
<213> Homo sapien

<400> 150

Met Gly Leu Thr Leu Leu Tyr Ser Ile Gly Glu Lys Asn Tyr Ile
1 5 10 15

Pro Thr Glu Lys Thr Glu Gly Glu Ala Ile Thr Thr Thr Lys Gln Ser
20 25 30

Val Thr Pro Arg Arg Glu Glu Met Gly Phe Pro Arg His Thr Pro His
35 40 45

Asn His Leu Gln Gln Pro Gln Pro Ser
50 55

<210> 151
<211> 28
<212> PRT
<213> Homo sapien

<400> 151

Met Phe Arg Gly Gln Ala Asp Ile Ile Thr Trp Cys Thr Phe Ser Ser
1 5 10 15

His Cys Leu Ala Lys Gly Ser Arg Ser Thr Ser Ser
20 25

<210> 152
<211> 13
<212> PRT
<213> Homo sapien

<400> 152

Met Ser Ser Gly Ala Gly Glu Asp Ser Gly Ala Gly Arg
1 5 10

<210> 153
<211> 87
<212> PRT
<213> Homo sapien

<400> 153

100

Met Gly Ala Leu Phe Pro Leu Pro Arg Tyr Ile Leu Thr Arg Leu Arg
 1 5 10 15

Ser Val Val Leu Ala Cys Gly Arg Val Glu Asn Gln Gly Ser Leu Lys
 20 25 30

Met Cys Gly Leu Tyr Thr Val Tyr Pro Gln Asn Ser Gly Asp Asn Ala
 35 40 45

Gly Glu Asn Asn His Val Glu Thr Lys Lys Cys His Ala Asn Lys Gly
 50 55 60

Gln Glu Pro Gly Lys Lys Gly Ser Arg Phe Val Cys Asp Val Ile Phe
 65 70 75 80

His Met Ala Ser Ser Pro His
 85

<210> 154
 <211> 57
 <212> PRT
 <213> Homo sapien

<400> 154

Met Ser Tyr Val Pro Cys Phe Tyr Ser Asn Val Asn Ser Ser Asn Phe
 1 5 10 15

Phe Ala Phe Phe Leu Leu Val Asn Val Cys Val Ile Ser Phe Val Phe
 20 25 30

Ile Asp Val Thr Trp Phe Tyr Phe Phe Ile Leu Leu Gln Phe Thr Ser
 35 40 45

Ile Ser Gly Thr Leu Phe Ala Ala Lys
 50 55

<210> 155
 <211> 115
 <212> PRT
 <213> Homo sapien

<400> 155

Met Phe Val Gly Gly Glu Leu Leu Arg Pro Glu Glu Pro Gln Phe His
 1 5 10 15

Pro Thr Gly Thr His Thr Tyr Ser Thr Gln Glu Val Pro Pro Lys Arg

101

20 25 30
 Phe Phe Phe Phe Phe Phe Phe Cys Asn Leu Pro Lys Ser Asn His
 35 40 45
 Pro Thr Phe Leu Glu Ile Leu Lys Thr Pro Lys Arg Lys Ile Ile Ser
 50 55 60
 Asn Asn Ser Thr Pro Thr Ser Lys Ala Phe Val Met Arg His Ser Gln
 65 70 75 80
 Ser Ile Phe Phe Phe Phe Phe Phe Leu Val Arg Val Ser Val Thr Gln
 85 90 95
 Ala Gly Ile Gln Trp Cys Asp Leu Ser Ser Pro Gln Pro Pro Pro Pro
 100 105 110
 Arg Phe Lys
 115
 <210> 156
 <211> 67
 <212> PRT
 <213> Homo sapien
 <400> 156
 Met Cys Val Tyr Ile Ser Pro Gly Ser Thr His Lys Phe Ser His Thr
 1 5 10 15
 Pro His Thr His Ile Ile Leu Gly Arg Ala Thr Gln Asn Ala Lys Lys
 20 25 30
 Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys Met Lys Lys Lys
 35 40 45
 Lys Lys Lys Lys Lys Lys Glu Lys Ile Lys Glu Asn Gln Arg Gln Thr
 50 55 60
 Glu Lys Thr
 65
 <210> 157
 <211> 51
 <212> PRT
 <213> Homo sapien

This page is not part of
the pamphlet!

WO 02-068645

4/4

Date: 06 sep 2002

Destination: Agent

Address:

102

<400> 157

Met His Ile Tyr Leu Val Arg Ile Pro Phe Gly Leu Leu Asn Arg Leu
 1 5 10 15

Thr Leu Glu Phe Ala Gln Asp Thr Glu Ala Asn Leu Ser Ala Gly Lys
 20 25 30

Asn Pro Asp Ala Pro His Ile Leu Arg Glu Pro His Met Ser Cys Ser
 35 40 45

Tyr Cys Cys
 50

<210> 158

<211> 135

<212> PRT

<213> Homo sapien

<400> 158

Met Phe Phe Val Arg Ala Cys Ile Leu Phe Tyr Thr Gln Tyr Leu Ser
 1 5 10 15

Phe Glu Trp His Leu Gln Tyr Ala Ala Pro Thr Pro Ser Phe Cys Ser
 20 25 30

Leu Arg His Leu Leu Cys Ser Cys Leu Pro His Phe Tyr Cys Leu Val
 35 40 45

Val Cys Leu Leu Pro Ala Ser Leu Ser Val Leu Pro Pro Phe Leu Phe
 50 55 60

Leu Pro Leu Leu Ala Leu Asp Thr Leu Phe Ala Val Thr Arg Lys Cys
 65 70 75 80

Leu Cys Gly Gly Lys Phe Val Glu Ser Arg Glu Arg Tyr Thr His Ile
 85 90 95

Val Thr His Thr Arg Gly Thr His Ser Tyr Trp Arg Pro Gln Arg Val
 100 105 110

Phe Thr Pro Gln Arg Leu Phe Ser Leu Phe Ile Ile Ser Pro Arg Glu
 115 120 125

Lys Asn Tyr Lys Glu Val Ile
 130 135

103

<210> 159
<211> 102
<212> PRT
<213> Homo sapien

<400> 159

Met Arg Val Val Pro Glu Met Val His Val Val Gln Val Ile Cys Leu
1 5 10 15

Leu Met Phe Val Ser Leu Phe Ile His Gly Val Asp Trp Arg Glu Gly
20 25 30

Thr Lys Ser Ile Cys Leu Tyr Ile Arg Thr Ser Val Val Arg Cys Ile
35 40 45

Phe His Val Thr Ser Leu Leu Glu Asp Gln Thr Pro Tyr Val Leu Gln
50 55 60

Tyr Ala Leu Pro Met Ala Val Leu Arg Arg Lys Leu Arg Leu Phe Cys
65 70 75 80

Phe Asn Arg Gly Trp Cys Thr Trp Leu Ser Lys Tyr Ser Val Lys Ser
85 90 95

Ser Ile Ser Glu Gly Asn
100

<210> 160
<211> 21
<212> PRT
<213> Homo sapien

<400> 160

Met Ser Val Leu Ser Val Ala Glu Leu Ser Val Ser Trp His Ser Cys
1 5 10 15

Ala Cys Val Lys Leu
20

<210> 161
<211> 16
<212> PRT
<213> Homo sapien

<400> 161

104

Met Thr Thr Ser Val Val Asn Phe Arg Asn Tyr Phe Phe Thr Ser Val
1 5 10 15

<210> 162
<211> 85
<212> PRT
<213> Homo sapien

<400> 162

Met Arg Gly Phe Leu Phe Pro Asp Gly Ile Gln Gly Ala Thr Ser Phe
1 5 10 15

Phe Leu Pro Gly Lys Lys Arg Tyr Thr Cys Cys Leu Asp Ser Ser Pro
20 25 30

His Phe Pro Pro Val Leu His His Gly Pro Leu Asn Phe Leu Phe Val
35 40 45

Leu Leu Pro Pro Ser Asn Asn His Glu Asn Asn Leu Gly Glu Val Phe
50 55 60

Gln Ile Met Lys Lys Lys Gln Lys Lys Gln Lys Asn Asn Gln Arg Gly
65 70 75 80

Asp Leu Gly Arg Asp
85

<210> 163
<211> 40
<212> PRT
<213> Homo sapien

<400> 163

Met Tyr Leu Thr Leu Ser Phe Ser Val Met Tyr Asn Cys His Phe Leu
1 5 10 15

Ile Leu Tyr Ile Met Tyr Leu Phe Asp Ile Arg Phe Asn Asn Tyr Ile
20 25 30

Asn Phe Ile His Ser Leu Phe Glu
35 40

<210> 164
<211> 33
<212> PRT
<213> Homo sapien

105

<400> 164

Met Ser Pro Gln Gln Thr Ile Leu Arg Val Ile Pro Glu Gln Lys Ser
 1 5 10 15

Thr Thr Thr Gln Leu Thr Leu Ile Leu Ser Leu Thr Lys Ser Ile Thr
 20 25 30

Leu

<210> 165

<211> 46

<212> PRT

<213> Homo sapien

<400> 165

Met Glu Leu Pro Phe Asn Lys Glu Ile Leu Pro Lys Gln Lys Lys Lys
 1 5 10 15

Lys Lys Lys Lys Lys Gly Trp Gly Ser Trp Pro Ala Val Pro Val Leu
 20 25 30

Asn Trp Phe Ser Gly Pro Lys Phe Pro Lys Ile Arg Glu Gln
 35 40 45

<210> 166

<211> 24

<212> PRT

<213> Homo sapien

<400> 166

Met Ala Ile Val Pro Leu Asp His Ala Ser Ser Gly Ala Ser Cys Asp
 1 5 10 15

Gly Leu Val Ala Ala Arg Tyr Asn
 20

<210> 167

<211> 75

<212> PRT

<213> Homo sapien

<400> 167

Met Thr Thr Tyr Ala Ile Gly Cys Glu Asp Glu Ala Ile Ala Ala Lys
 1 5 10 15

106

Pro Gly Val Ser Asn Asp Asn Glu Arg Arg Pro Cys Thr Ile Val Leu
 20 25 30

Glu Leu Arg Arg Glu Pro Leu Ser Leu Ser Ser Pro Ile Ser Lys Ala
 35 40 45

Leu Pro Val Asn Gln Glu Thr Ala Cys Thr Thr Cys Val Glu Gln Ser
 50 55 60

Leu Ser Leu Leu His Asp Ala Pro Met Leu Val
 65 70 75

<210> 168
 <211> 91
 <212> PRT
 <213> Homo sapien

<400> 168

Met Leu Cys His His Val Ile Arg Tyr Asn Leu His Phe Ser Val Leu
 1 5 10 15

Thr Ser His Pro Ile Tyr Thr Val Leu Tyr Ala His Lys Cys Ile Gly
 20 25 30

Gly Arg His Gln Phe Val Met Ala His Val Ser His Asn Met Lys Tyr
 35 40 45

Leu Glu Glu Leu Leu Tyr Val Gly Glu Cys Pro Tyr Val Gly Val Asn
 50 55 60

Val Ser Met Tyr Phe Leu Arg Val Ala Arg Pro Thr Cys Leu Leu Cys
 65 70 75 80

Phe Thr Tyr Asp Phe Tyr Thr Arg Ala Arg Ala
 85 90

<210> 169
 <211> 211
 <212> PRT
 <213> Homo sapien

<400> 169

Met Ala Ala Glu Ala Thr Thr Glu Arg Arg Arg Arg Glu Ser Glu Glu
 1 5 10 15

Thr Arg Arg Arg Glu Arg Ala Arg Arg Arg Asn Glu Arg Arg Lys Arg

107

20

25

30

Gly Ala Glu Ala Glu Arg Gly Asp Arg Thr Ala Arg Glu Glu Ser Glu
 35 40 45

Ala Pro Asn Gly Glu Arg Asn Asn Glu Arg Glu Thr Asp Glu Thr Arg
 50 55 60

Thr Gln Arg Arg Arg Arg Thr Thr His Arg Gln Arg Arg Glu Lys Thr
 65 70 75 80

Ser Arg Glu Ala His His Gly Gln Ser Ala Glu Ala Gln Pro Gln Glu
 85 90 95

Thr Thr Thr Gly Pro Arg Glu Gln Arg Arg Gln Met Arg Ala Glu Ala
 100 105 110

Thr Arg Thr Thr Val Lys Asp Gln Asp Glu Thr Ser Ser Lys Glu Lys
 115 120 125

Arg Arg Met Arg Thr His Asn Ile Lys Ile Arg Gln Thr Arg Ser Gly
 130 135 140

Thr His Asp Ala Arg Gln Arg Glu Glu Arg His Thr Thr Asn Lys His
 145 150 155 160

Ala Arg Ser Arg Gly Gln His Glu Arg Lys Gln Pro Glu Gln Lys Gln
 165 170 175

Glu Ser Ala Gly Lys Arg Arg Gly Asp Ser Ser Asn Arg Arg Ala Thr
 180 185 190

Gln Arg Arg Lys Arg Leu Glu Lys Glu Lys Thr Gln Lys Thr Arg His
 195 200 205

Gly Arg His
 210

<210> 170
 <211> 82
 <212> PRT
 <213> Homo sapien

<400> 170

Met Phe Ile Ser Val Phe His Val Trp Phe Val Ala Val Val Val Gly

108

1 5 10 15
 Glu Ile Gly Ser Arg Gly Lys His Asn Phe Tyr Thr Pro Arg Asn Gln
 20 25 30
 Arg Leu Ala Pro Arg Ser Phe Pro Arg Pro Ala Ser Leu Val Tyr Thr
 35 40 45
 Arg Asn Ile Ser Cys Ser Phe Ser Pro Gln Arg Thr His Gly Arg Asp
 50 55 60
 Thr Gly Ser Leu Gly Pro His Val Met Lys Arg Tyr Trp Ala Pro Pro
 65 70 75 80
 Thr Ala

 <210> 171
 <211> 153
 <212> PRT
 <213> Homo sapien

 <400> 171
 Met Ser Leu Ala Asp Gly His Ser Trp Arg Pro Gln Phe Met Phe Asn
 1 5 10 15
 Arg Asn Ser Leu Arg Asn Ile Leu Arg Leu Pro His Pro Leu Val Val
 20 25 30
 Leu Pro Ser Phe Leu Pro Ser Leu Arg Val Lys Gly Pro Arg Gly Pro
 35 40 45
 Phe Trp Val Leu Leu Trp Lys Ala Arg Asp Val Ser Val Phe His Arg
 50 55 60
 Thr Ala Trp Arg Pro Lys His Pro Gly Ala Pro Ile Gly Arg Gly Ser
 65 70 75 80
 Pro Gly Gly Val Thr Val Trp Phe Tyr Arg Arg Ser Pro Lys Leu Pro
 85 90 95
 Pro Pro His His Cys Gln Gln Gln Lys Val Gly Pro Leu Gly Ala Gly
 100 105 110
 Ala Thr Met Leu Asn Thr Gly Ser Ser Arg Glu His Ala Ala Gln Ala

109

115

120

125

Thr Lys Ala Gly Arg Ser Lys Thr Gln Ala His Thr Lys Asn Glu Ile
 130 135 140

Ser Lys Gln Ala Thr Glu Gln Ala Ser
 145 150

<210> 172
 <211> 32
 <212> PRT
 <213> Homo sapien

<400> 172

Met Gln Pro Arg Gly Ser Thr Asp Asn Arg Ile Leu Lys Lys Val Ala
 1 5 10 15

Ala Pro Pro Val Ile Ile Asn Asn Leu Ile Lys Phe Thr Glu Leu Tyr
 20 25 30

<210> 173
 <211> 48
 <212> PRT
 <213> Homo sapien

<400> 173

Met Ser Val Gly Trp Asp Cys Ser Gln Val Tyr Ile Thr Lys Arg Ile
 1 5 10 15

Gly Ala Thr His Val Gly Phe Met Phe Cys Asp Val Leu Ser Ile Cys
 20 25 30

Val Asn Ala Phe His Met Val Ser Gly Leu Glu Cys Tyr Gly Pro Leu
 35 40 45

<210> 174
 <211> 17
 <212> PRT
 <213> Homo sapien

<400> 174

Met Lys Thr Gln Glu Lys Arg Met Val Asn Lys Glu Asp Pro Asn Tyr
 1 5 10 15

Leu

110

<210> 175
 <211> 132
 <212> PRT
 <213> Homo sapien

<400> 175

Val Val Met Thr Leu Asn Glu His Ala Ala Phe Lys His Leu Phe Asn
 1 5 10 15

Lys Ala His Leu Ala Pro Pro Leu Ile His Leu Thr Leu Ser Gly His
 20 25 30

Ser Thr Cys Phe Arg Glu His Arg Val Gly Asp Lys Val Thr Asp Gln
 35 40 45

Gln Asp Pro Lys Ala Glu Glu Phe Phe Leu Val Gln Asn Lys Met Lys
 50 55 60

Ser Leu Pro Cys Leu Leu Leu Ser Thr Glu Thr Arg Gln Pro Ser Asp
 65 70 75 80

Phe Ser Ile Phe Ser Pro Leu Phe Pro Leu Phe Tyr Ser Thr Lys Pro
 85 90 95

Pro Leu Ser Ser Trp Pro Val Leu Asn Glu Leu Leu Gly Thr Pro Pro
 100 105 110

Arg Arg Gly Gly Gly Arg Ala Glu Gly Leu Leu Thr Ser Gln Gly Leu
 115 120 125

Leu Thr Ser Gln
 130

<210> 176
 <211> 114
 <212> PRT
 <213> Homo sapien

<400> 176

Met Ile Glu Leu Leu Ser Ser Ser Val Tyr His Glu Gly Pro Pro His
 1 5 10 15

Ala Val Phe Gly Ala Pro Val Leu Pro Pro Ser Val Ser Cys Ile Val
 20 25 30

111

Cys Thr Thr Pro Pro Gln Leu Gly Gly Pro Pro Pro Pro Pro Leu
 35 40 45

Val His Ala Thr Phe Pro Pro Pro Phe Pro Arg Thr Thr Pro Pro Phe
 50 55 60

Phe Thr Pro Pro Pro Pro Pro Phe Leu Leu Phe Pro Pro Pro Pro Pro
 65 70 75 80

Pro Pro Arg Val Phe Phe Phe Lys Lys Lys Lys Lys Lys Lys Lys Lys
 85 90 95

Gln Lys Lys Lys Lys Lys Lys Lys Lys Gly Gly Gly Thr Cys Pro Ala
 100 105 110

Ala Ala

<210> 177
 <211> 43
 <212> PRT
 <213> Homo sapien

<400> 177

Met Pro Tyr Leu Arg Leu Trp Lys Asn Gly Val Tyr Ser Pro Cys Asn
 1 5 10 15

Phe Leu Gly Glu Lys Lys Pro Phe Pro Met Asp Leu Lys Lys Lys Lys
 20 25 30

Lys Lys Lys Lys Lys Asn Leu Ala Ala Thr Thr
 35 40

<210> 178
 <211> 213
 <212> PRT
 <213> Homo sapien

<400> 178

Met Thr Ser Asp Glu Ala Thr Thr Glu Thr Arg Pro Ala Arg Glu Ala
 1 5 10 15

Glu Lys Gly Ala Glu Lys Gln Lys Ala Thr Glu Lys Gly Lys Thr Lys
 20 25 30

Lys Thr Ser Thr Ser Tyr Arg Arg Ser Gln Arg Met Arg Lys Glu Arg

112

35

40

45

Arg Arg Arg Lys His Glu Ala Thr Arg Arg Arg Thr Gly Glu Glu Arg
 50 55 60

Glu Asn Arg Gly Arg Arg Arg Glu Gln Arg Arg Arg Arg Thr Lys Val
 65 70 75 80

Gly Ser Gln Glu Glu Thr Lys Arg Glu Val Gln Thr Glu Gln Gly Arg
 85 90 95

Lys Arg Pro Lys Gly Gln Lys Lys Glu Thr Gln Arg Arg Lys Lys Arg
 100 105 110

Arg Lys Lys Lys Ser Gln Arg Arg Arg Thr Gly Lys Arg Lys Gln Glu
 115 120 125

Glu Lys Thr Thr Gln Arg Glu Arg Arg Glu Lys Asp Lys Arg Ser Arg
 130 135 140

Arg Glu Trp Lys Tyr Ala Glu Glu Glu Glu Thr Asp Asn Glu Glu Arg
 145 150 155 160

Arg Arg Lys Lys Arg Lys Arg Gln Gln Lys Lys Arg Glu Lys Lys Arg
 165 170 175

Arg Ser Lys Lys Ser Arg Ser Lys Asn Glu Ala Asp Lys Glu Arg Ala
 180 185 190

Glu Thr Thr Arg Arg Glu Glu Arg Glu Arg Glu Thr Glu Glu Glu Lys
 195 200 205

Thr Arg Asn Arg Ser
 210

<210> 179

<211> 434

<212> PRT

<213> Homo sapien

<400> 179

Met Ser Ala Asp Ala Ala Ala Gly Ala Pro Leu Pro Arg Leu Cys Cys
 1 5 10 15

Leu Glu Lys Gly Pro Asn Gly Tyr Gly Phe His Leu His Gly Glu Lys

113

20

25

30

Gly Lys Leu Gly Gln Tyr Ile Arg Leu Val Glu Pro Gly Ser Pro Ala
 35 40 45

Glu Lys Ala Gly Leu Leu Ala Gly Asp Arg Leu Val Glu Val Asn Gly
 50 55 60

Glu Asn Val Glu Lys Glu Thr His Gln Gln Val Val Ser Arg Ile Arg
 65 70 75 80

Ala Ala Leu Asn Ala Val Arg Leu Leu Val Val Asp Pro Glu Thr Asp
 85 90 95

Glu Gln Leu Gln Lys Leu Gly Val Gln Val Arg Glu Glu Leu Leu Arg
 100 105 110

Ala Gln Glu Ala Pro Gly Gln Ala Glu Pro Pro Ala Ala Ala Glu Val
 115 120 125

Gln Gly Ala Gly Asn Glu Asn Glu Pro Arg Glu Ala Asp Lys Ser His
 130 135 140

Pro Glu Gln Leu Ser Leu Val Ala Val Ser Asp Gly Ser Val Arg Gly
 145 150 155 160

Ala Thr Arg Ser Leu Leu Asp Arg Glu Arg Ala Gln Phe Gly Ile Lys
 165 170 175

Arg Gln Asn Pro Ala Leu Pro Gln Leu Gly Gly Glu Gly Pro Arg Ala
 180 185 190

Met Val Ala Glu Leu Gly Gln Arg Glu Leu Arg Pro Arg Leu Cys Thr
 195 200 205

Met Lys Lys Gly Pro Ser Gly Tyr Gly Phe Asn Leu His Ser Asp Lys
 210 215 220

Ser Lys Pro Gly Gln Phe Ile Arg Ser Val Asp Pro Asp Ser Pro Ala
 225 230 235 240

Glu Ala Ser Gly Leu Arg Ala Gln Asp Arg Ile Val Glu Val Asn Gly
 245 250 255

114

Val Cys Met Glu Gly Lys Gln His Gly Asp Val Val Ser Ala Ile Arg
 260 265 270

Ala Gly Gly Asp Glu Thr Lys Leu Leu Val Val Asp Arg Glu Thr Asp
 275 280 285

Glu Phe Phe Lys Lys Cys Arg Val Ile Pro Ser Gln Glu His Leu Asn
 290 295 300

Gly Pro Leu Pro Val Pro Phe Thr Asn Gly Glu Ile His Lys Asp Pro
 305 310 315 320

Leu Thr Pro Ser Ser Asp Asn Pro Gln Pro Ser Pro Leu Cys Gln Glu
 325 330 335

Asn Ser Arg Glu Ala Leu Ala Glu Ala Ala Leu Glu Ser Pro Arg Pro
 340 345 350

Ala Leu Val Arg Ser Ala Ser Ser Asp Thr Ser Glu Glu Leu Asn Ser
 355 360 365

Gln Asp Ser Pro Pro Lys Gln Asp Ser Thr Ala Pro Ser Ser Thr Ser
 370 375 380

Ser Ser Asp Pro Ile Leu Asp Phe Asn Ile Ser Leu Ala Met Ala Lys
 385 390 395 400

Glu Arg Ala His Gln Lys Arg Ser Ser Lys Arg Ala Pro Gln Met Asp
 405 410 415

Trp Ser Lys Lys Asn Glu Leu Phe Ser Asn Leu Asn Glu Leu Phe Ser
 420 425 430

Asn Leu

<210> 180
 <211> 49
 <212> PRT
 <213> Homo sapien

<400> 180

Met Gly Ser Cys Ser Val Ala Gln Val Gly Val Met Trp His Asp Leu
 1 5 10 15

115

Gly Ser Leu Gln Pro Leu Pro Pro Gly Phe Lys Gln Phe Ser Cys Leu
 20 25 30

Ser Leu Leu Ser Ser Trp Asp Tyr Arg Arg Thr Cys Pro Gly Gly Arg
 35 40 45

Ser

<210> 181
 <211> 59
 <212> PRT
 <213> Homo sapien

<400> 181

Phe Phe Phe Leu Phe Val Cys Leu Phe Glu Met Gly Ser Cys Ser Val
 1 5 10 15

Ala Gln Val Gly Val Met Trp His Asp Leu Gly Ser Leu Gln Pro Leu
 20 25 30

Pro Pro Gly Phe Lys Gln Phe Ser Cys Leu Ser Leu Leu Ser Ser Trp
 35 40 45

Asp Tyr Arg Cys Glu Pro Gln Arg Leu Ala Arg
 50 55

<210> 182
 <211> 193
 <212> PRT
 <213> Homo sapien

<400> 182

Met Ser Tyr Ser Phe Ala Ser Ser Val Val Leu Val Asp Ser Leu Thr
 1 5 10 15

Ser Phe Leu Gly Pro Phe Thr Phe Ser Leu Leu Ala Thr Ser Arg Ile
 20 25 30

Leu His Leu Tyr Leu Ala Pro Arg Val Arg Leu Ser Cys Ser Ser Leu
 35 40 45

Ser Pro Phe Ala Cys Leu Leu Cys Ser Leu Leu Trp Val Arg Val Ser
 50 55 60

Ser Ser Ser Thr Arg Ser Ile Cys Ser Leu Ser Val Phe Cys Val Cys

```
<210> 183
<211> 56
<212> PRT
<213> Homo sapien

<400> 183
```

```

Met Trp Cys Arg Cys Val Cys Leu Asn Tyr Cys Gln Cys Val Pro Pro
1          5          10          15

Ser Trp Thr Phe Leu Pro Ser Leu Met His Val Gln Tyr Asp Ser His
          20          25          30

Glu Asn Asp Glu Pro Cys His Glu Val Leu Ile Ala Asn Glu Glu Arg
          35          40          45

Leu His Arg Lys Asn Met Lys Lys
          50          55

```

<210> 184

117

<211> 105
 <212> PRT
 <213> Homo sapien

<400> 184

Met Pro Tyr Gly Val Thr Gln Phe Lys Leu Thr Arg Ile Val Ser Ala
 1 5 10 15

Ile Gly Trp Glu Leu Thr Thr Cys Asp Pro Ser Tyr Tyr Thr Pro Val
 20 25 30

Leu Thr Leu Ser Leu Leu Lys Phe Cys Ala Leu Glu His Ile His Lys
 35 40 45

Asn Asn Arg Ala Arg Ala Leu Gln Gly Asn His Thr Pro Pro Asn Ser
 50 55 60

Lys Leu Arg Asn Thr His Ile Ser Arg Glu Ala Gln Arg Gly Tyr Lys
 65 70 75 80

Glu Tyr Cys Ala Arg Gln Arg Asn Pro Gln Thr Pro His Pro Arg Ala
 85 90 95

Gln Pro Gly Thr Gln Asn Ser Lys Asn
 100 105

<210> 185
 <211> 38
 <212> PRT
 <213> Homo sapien

<400> 185

Met Ile Val Arg Gly Glu Val His Thr Leu Met His Leu Glu Leu Tyr
 1 5 10 15

Cys Ile Ile Arg Thr Thr Ser Asp Thr Ser Phe Phe Phe Phe Phe Phe
 20 25 30

Phe Phe Pro Tyr Cys Asn
 35

<210> 186
 <211> 77
 <212> PRT
 <213> Homo sapien

<400> 186

118

Met Val Thr Gly Cys Leu Leu Arg Gln Cys Ala Asp Arg Cys Gln Val
 1 5 10 15

Asn Ser Thr Ala His Phe Trp Leu Asn Phe Leu Gln Leu Ser Ser Val
 20 25 30

Arg Ser Lys Val His Leu Gln Pro Ser Leu Arg Ala Leu Leu Phe Ser
 35 40 45

Ser Ser Val Arg Thr Cys Thr Gly Gln Pro Cys Pro Phe Gln Phe Ser
 50 55 60

Ala Ser Trp Leu Gly Ala His Arg Leu Leu Ser Asn His
 65 70 75

<210> 187
 <211> 13
 <212> PRT
 <213> Homo sapien

<400> 187

Met Leu Phe Pro Cys Val Lys Leu Val Tyr Ser Ala His
 1 5 10

<210> 188
 <211> 44
 <212> PRT
 <213> Homo sapien

<400> 188

Met Arg Arg Pro Ala Arg Leu Val Glu Arg Ala Val Cys Leu Val Leu
 1 5 10 15

Glu Phe Leu Phe Phe Ile Ser Phe Leu Ser Cys Asn Ser Tyr Phe Trp
 20 25 30

Phe Ala Trp Thr Val Leu His Thr Pro Ile Phe Leu
 35 40

<210> 189
 <211> 53
 <212> PRT
 <213> Homo sapien

<400> 189

Met Leu Leu Ser Lys Gly Thr Gly Thr Thr Leu Ile Phe Ile Asp Gly

1 5 10 15

Ser Phe Arg Gly Asn His Pro Phe Pro Thr Glu Pro Thr Pro Phe Lys
20 25 30

120

Leu Asn Ser Ser Leu Arg Leu Leu Gly Phe Ser Leu Ala Val Lys Ser
 35 40 45

Ser Gly Phe Leu Lys Asn Asp Gly Leu Pro Trp Lys
 50 55 60

<210> 192
 <211> 269
 <212> PRT
 <213> Homo sapien

<400> 192

Met Ala Ala Ser Gly Ser Gly Met Ser Gln Lys Thr Trp Glu Leu Ala
 1 5 10 15

Asn Asn Met Gln Glu Ala Gln Ser Ile Asp Glu Ile Tyr Lys Tyr Asp
 20 25 30

Lys Lys Gln Gln Gln Glu Ile Leu Ala Ala Lys Pro Trp Thr Lys Asp
 35 40 45

His His Tyr Phe Lys Tyr Cys Lys Ile Ser Ala Leu Ala Leu Leu Lys
 50 55 60

Met Val Met His Ala Arg Ser Gly Gly Asn Leu Glu Val Met Gly Leu
 65 70 75 80

Met Leu Gly Lys Val Asp Gly Glu Thr Met Ile Ile Met Asp Ser Phe
 85 90 95

Ala Cys Leu Trp Gln Gly Thr Glu Thr Arg Val Asn Ala Gln Ala Ala
 100 105 110

Ala Tyr Glu Tyr Met Ala Ala Tyr Ile Glu Asn Ala Lys Gln Val Gly
 115 120 125

Arg Leu Glu Asn Ala Ile Gly Trp Tyr His Ser His Pro Gly Tyr Gly
 130 135 140

Cys Trp Leu Ser Gly Ile Asp Val Ser Thr Gln Met Leu Asn Gln Gln
 145 150 155 160

Phe Gln Glu Pro Phe Val Ala Val Val Ile Asp Pro Thr Arg Thr Ile
 165 170 175

121

Ser Ala Gly Lys Val Asn Leu Gly Ala Phe Arg Thr Tyr Pro Lys Gly
 180 185 190

Tyr Lys Pro Pro Asp Glu Gly Pro Ser Glu Tyr Gln Thr Ile Pro Leu
 195 200 205

Asn Lys Ile Glu Asp Phe Gly Val His Cys Lys Gln Tyr Tyr Ala Leu
 210 215 220

Glu Val Ser Tyr Phe Lys Ser Ser Leu Asp Arg Lys Leu Leu Glu Leu
 225 230 235 240

Leu Trp Asn Lys Tyr Trp Val Asn Thr Leu Ser Ser Ser Ser Leu Leu
 245 250 255

Thr Asn Ala Asp Tyr Thr Thr Gly Gln Val Phe Asp Leu
 260 265

<210> 193
 <211> 146
 <212> PRT
 <213> Homo sapien

<400> 193

Met Trp Cys Ser Tyr Pro Tyr Cys Cys Ser Gly Phe Leu Leu Ser Tyr
 1 5 10 15

Thr Val Cys Thr His Gly Val Asn Ile Gly Cys Val Cys Cys Leu Ser
 20 25 30

Arg Trp Trp Leu Ser Leu Val Met Val Pro Val Pro Cys Val Val Val
 35 40 45

Phe Thr Ala Cys Trp Val Cys Val Trp Ser Ser Glu Pro His Leu Met
 50 55 60

Asp Met Trp Val Arg Pro Val Val His Phe Leu Ala Met Cys His Val
 65 70 75 80

Pro Arg Val Cys Ser Leu Phe Pro Leu Leu Val Cys Ala Cys Ser Phe
 85 90 95

Leu Phe Leu Leu Gly Ile Leu Ala Leu Cys Pro Pro Val Ala Leu Tyr
 100 105 110

122

Ser Leu Gly Val Cys Val Ser Pro Pro Val Ile Cys Ser Pro Ala Cys
 115 120 125

Glu Ile Trp Trp Val Cys Arg Ala Pro Ser Cys Ala Leu Tyr Pro Leu
 130 135 140

Arg Pro
 145

<210> 194
 <211> 141
 <212> PRT
 <213> Homo sapien

<400> 194

Met Cys Ala His Thr His Gly Ala Gly His Thr Ala Leu His Phe Gly
 1 5 10 15

Arg His Ala Gln Val Phe Ile Arg Arg Ala Arg Gly Leu Ser Ser Ser
 20 25 30

Arg Ile Thr His Ser Glu Ser Tyr Cys Leu Leu Pro Ser Leu His Thr
 35 40 45

Gln Gly Thr Pro Arg Ser Arg Gly Arg Pro Thr Arg Gly Val Ser Leu
 50 55 60

Ser Ser Arg Ala Leu Val Leu Arg Arg Glu Val Leu Gly Asp Thr His
 65 70 75 80

Thr His Thr Pro Glu Ser Gly Asp Thr Arg Tyr Arg Asp Cys Leu His
 85 90 95

Thr Lys Ile Phe Tyr Asn Ile Glu Ile Cys Gly Ser Arg Thr Gln His
 100 105 110

Ile Trp Ala Pro Ala His Thr Glu Thr Leu Ser Ser Leu Ser His Arg
 115 120 125

Ala Val Ala Pro Leu Leu His Arg Glu Ser Gly Glu Pro
 130 135 140

<210> 195
 <211> 95

123

<212> PRT

<213> Homo sapien

<400> 195

Met Ser Ser His Leu Thr Asn Ser Cys Val Phe Pro Lys Tyr Ser Ser
 1 5 10 15

Leu Phe Thr Gln Gly Leu Val Val Lys Ile Tyr Gln His Pro Gly Ile
 20 25 30

Lys Phe Ser Leu Trp Glu Ser Leu Phe His Lys Lys Trp Ala Pro Gly
 35 40 45

Phe Leu Thr Pro Leu Val Trp Lys Met Leu Trp Gly Glu Met Glu Lys
 50 55 60

Ser His Phe Leu Leu Tyr Leu Asn Ala Gly Gly Glu Thr Ser Trp Ala
 65 70 75 80

Asn Ser Arg Val Pro Val Val Gly Lys Trp Leu Ser Pro Pro Gln
 85 90 95

<210> 196

<211> 54

<212> PRT

<213> Homo sapien

<400> 196

Met Arg Thr Val Val Ile Pro Glu Gly Trp Gly Gly Asp Arg Leu Gly
 1 5 10 15

Glu Gly Phe Arg Lys Leu Ser Glu Asp Asp Cys Asn Gly Leu Asn Phe
 20 25 30

Gly Lys Val Trp Leu His Arg Cys Ile Cys Leu Gln Glu Leu Ser Lys
 35 40 45

Phe Ile Leu Lys Ile Cys
 50

<210> 197

<211> 240

<212> PRT

<213> Homo sapien

<400> 197

124

Met Pro Pro Leu Leu Phe Glu Val Ser Ser Leu Glu Asn Ala Phe Gln
 1 5 10 15

Ile Gly Gly His Pro Trp His Tyr Ile Val Thr Pro Asn Lys Lys Lys
 20 25 30

Gln Lys Gly Val Phe His Ile Cys Ala Leu Lys Asp Asn Ser Leu Ala
 35 40 45

Lys Asn Gly Ile Gln Glu Met Asp Cys Cys Ser Leu Glu Ser Asp Trp
 50 55 60

Ile Tyr Phe His Pro Asp Ala Ser Gly Arg Ile Ile His Val Gly Pro
 65 70 75 80

Asn Gln Val Lys Val Leu Lys Leu Thr Glu Ile Glu Asn Asn Ser Ser
 85 90 95

Gln His Gln Ile Ser Glu Asp Phe Val Ile Leu Ala Asn Arg Glu Asn
 100 105 110

His Lys Asn Glu Asn Val Leu Thr Val Thr Ala Ser Gly Arg Val Val
 115 120 125

Lys Lys Ser Phe Asn Leu Leu Asp Asp Asp Pro Glu Gln Glu Thr Phe
 130 135 140

Lys Ile Val Asp Tyr Glu Asp Glu Leu Asp Leu Leu Ser Val Val Ala
 145 150 155 160

Val Thr Gln Ile Asp Ala Glu Gly Lys Ala His Leu Asp Phe His Cys
 165 170 175

Asn Glu Tyr Gly Thr Leu Leu Lys Ser Ile Pro Leu Val Glu Ser Trp
 180 185 190

Asp Val Thr Tyr Ser His Glu Val Tyr Phe Asp Arg Asp Leu Val Leu
 195 200 205

His Ile Glu Gln Lys Pro Asn Arg Val Phe Ser Cys Tyr Val Tyr Gln
 210 215 220

Met Ile Cys Asp Thr Gly Glu Glu Glu Glu Thr Ile Asn Arg Ser Cys
 225 230 235 240

125

<210> 198
 <211> 31
 <212> PRT
 <213> Homo sapien

<400> 198

Met Ile Pro Gln Leu Gly Glu Ser Val Leu Ile His Cys Pro Asn Gly
 1 5 10 15

Pro Pro Leu Pro His Val Ser Pro Pro Ser Ser Asn Pro Ser Tyr
 20 25 30

<210> 199
 <211> 62
 <212> PRT
 <213> Homo sapien

<400> 199

Met Pro Ala Pro Leu Gly Gly Arg Gly Gly Trp Ser Pro Pro Arg Ser
 1 5 10 15

Arg Ser Ser Arg Gln Arg Leu Ala Asp Met Ala Lys Pro Arg Leu Tyr
 20 25 30

Tyr Lys Lys Asn Thr Lys Arg Leu Asp Trp Val Trp Trp Cys Val Pro
 35 40 45

Ile Ile Pro Ala Thr Gln Glu Ala Glu Ala Gly Glu Phe Phe
 50 55 60

<210> 200
 <211> 245
 <212> PRT
 <213> Homo sapien

<400> 200

Met Gly Arg Ser Cys Val Val Cys Phe Val Cys Leu Phe Phe Ser Phe
 1 5 10 15

Val Phe Arg Leu Ser Ser Arg Ala Val Ala Ala Leu Arg Phe Ser Val
 20 25 30

Cys Val Val Arg Arg Val Arg Leu Ala Ala Ser Ser Phe Val Leu Arg
 35 40 45

Arg Ser Ala Leu Ser Leu Ser Ser Val Ser Ser Leu Val Ser Pro Ala

126

50

55

60

Leu Leu Pro Leu Arg Ser Leu Ser Ser Ser Ser Phe Leu Ser Pro Phe
 65 70 75 80

Val Ala Pro Cys Leu Ser Val Cys Phe Val Pro Val Leu Val Cys Leu
 85 90 95

Ser Ser Ala Phe Ala Ser Leu Ser Arg Ser Cys Ser Phe Leu Leu Ser
 100 105 110

Val Arg Phe Ala Phe Ser Val Ser Arg Val Gly Leu Phe Cys Val Leu
 115 120 125

Phe Leu Leu Cys Leu Ala Arg Leu Ser Ser Val Phe Ala Ser Cys Ser
 130 135 140

Gly Phe Ser Leu Leu Phe Phe Phe Leu Leu Phe Phe Phe Phe Cys Phe
 145 150 155 160

Leu Ser Leu Cys Leu Ser Phe Phe Phe Ser Phe Leu Phe Phe Pro Ser
 165 170 175

Trp Cys Leu Phe Ser Phe Leu Phe Phe Ala Phe Ser Ser Ile Cys Phe
 180 185 190

Cys Leu Leu Trp Asp Asn Phe Leu Phe Val Phe Leu Ala Ile Phe Ser
 195 200 205

Ser Val Phe Ser Ser Leu His Cys Val Phe Leu Phe Ser Ser Phe Val
 210 215 220

Pro Pro Leu Tyr Phe Val Ile Phe Ser Phe Ala Leu Trp Tyr Ser Cys
 225 230 235 240

Trp Arg Pro Gly Val
 245

<210> 201
 <211> 144
 <212> PRT
 <213> Homo sapien

<400> 201

Glu Gln Met Ser Cys Gln Trp Glu Phe Lys Cys Gln His Gly Glu Glu

1				5						10					15		
Glu	Cys	Lys	Phe	Asn	Lys	Val	Glu	Ala	Cys	Val	Leu	Asp	Glu	Leu	Asp		
			20					25					30				
Met	Glu	Leu	Ala	Phe	Leu	Thr	Ile	Val	Cys	Met	Glu	Glu	Phe	Glu	Asp		
		35					40					45					
Met	Glu	Arg	Ser	Leu	Pro	Leu	Cys	Leu	Gln	Leu	Tyr	Ala	Pro	Gly	Leu		
		50				55					60						
Ser	Pro	Asp	Thr	Ile	Met	Glu	Cys	Ala	Met	Gly	Asp	Arg	Gly	Met	Gln		
65					70					75					80		
Leu	Met	His	Ala	Asn	Ala	Gln	Arg	Thr	Asp	Ala	Leu	Gln	Pro	Pro	His		
				85					90					95			
Glu	Tyr	Val	Pro	Trp	Val	Thr	Val	Asn	Gly	Lys	Pro	Leu	Glu	Asp	Gln		
			100					105					110				
Thr	Gln	Leu	Leu	Thr	Leu	Val	Cys	Gln	Leu	Tyr	Gln	Gly	Lys	Lys	Pro		
		115					120					125					
Asp	Val	Cys	Pro	Ser	Ser	Thr	Ser	Ser	Leu	Arg	Ser	Val	Cys	Phe	Lys		
		130				135					140						
<210>	202																
<211>	76																
<212>	PRT																
<213>	Homo sapien																
<400>	202																
Met	Pro	Ser	Asp	Arg	Met	His	Leu	Phe	Ile	Leu	Lys	Met	Ala	Ser	Leu		
1				5					10					15			
Arg	His	Pro	Thr	Gly	Gln	Pro	Cys	Lys	Leu	Lys	Ser	Gln	Gly	Ala	His		
			20					25					30				
Cys	Thr	Gln	Leu	Ser	His	Ala	Leu	Thr	Thr	Ala	Ser	Leu	Gln	Leu	Leu		
		35					40					45					
Thr	Leu	Gly	Tyr	Asn	Ser	Ser	Asn	Ile	Asn	Gly	Phe	Ser	Leu	Gln	His		
	50					55					60						
Cys	Thr	Leu	Gln	Asn	Ile	Glu	Gln	Gly	Phe	Ser	Leu						

128

65

70

75

<210> 203
 <211> 60
 <212> PRT
 <213> Homo sapien

<400> 203

Asp Ala Lys Glu Asp His Glu Arg Thr His Gln Met Val Leu Leu Arg
 1 5 10 15

Lys Leu Cys Leu Pro Met Leu Cys Phe Leu Leu His Thr Ile Leu His
 20 25 30

Ser Thr Gly Gln Tyr Gln Glu Cys Leu Gln Leu Ala Asp Met Val Ser
 35 40 45

Ser Glu Gly His Lys Leu Tyr Leu Val Ser Ser Arg
 50 55 60

<210> 204
 <211> 96
 <212> PRT
 <213> Homo sapien

<400> 204

Met Cys Leu Val Ser Phe Val Val Phe Ile Phe Leu Ser Asn Thr Pro
 1 5 10 15

Gly Pro Phe Phe Ser Phe Ser Leu Gly Leu Phe Ser Phe Ala Phe Leu
 20 25 30

Phe Leu Gln Leu Phe Phe Phe Leu Val Leu Phe Ser Phe Leu Ile Phe
 35 40 45

Leu Leu Val Phe Ser Val Phe Ser Leu Leu Asp Phe Tyr Phe Tyr Met
 50 55 60

Phe Val Phe Ser Phe Phe Ser Leu Leu Ser Leu Phe Ser Phe Leu Leu
 65 70 75 80

Phe Phe Tyr Val Val Val Leu Ser Trp Ile Leu Asp Trp Ile Phe Arg
 85 90 95

<210> 205
 <211> 34

129

<212> PRT
 <213> Homo sapien

<400> 205

Met Met Asp Asp Thr Leu Pro Gly Thr Leu Val His Tyr Ser Gln Cys
 1 5 10 15

Ser Ser Ser Ala Tyr Asn Ser Cys Leu Pro Val Asp Ser Thr Asn Glu
 20 25 30

Ser Gly

<210> 206
 <211> 42
 <212> PRT
 <213> Homo sapien

<400> 206

Met Pro Val Val Pro Ala Ile Trp Glu Ala Lys Glu Asp Arg Leu Ser
 1 5 10 15

Ser Gly Asp Arg Gly Cys Ser Trp Ala Glu Ile Ala Pro Gln Pro Ser
 20 25 30

Ser Leu Val Lys Arg Glu Arg Phe His Leu
 35 40

<210> 207
 <211> 111
 <212> PRT
 <213> Homo sapien

<400> 207

Leu Phe Val Tyr Ala Arg Trp Asn Leu Ser Leu Leu Thr Arg Leu Glu
 1 5 10 15

Gly Cys Gly Ala Ile Ser Ala Gln Cys Asn Leu Tyr Leu Leu Ser Ser
 20 25 30

Ser Asp Pro Ser Leu Ala Ser Gln Ile Ala Gly Thr Thr Gly Met Cys
 35 40 45

His His Val Gln Leu Ile Leu Tyr Phe Ala Ala Arg Arg Phe Tyr His
 50 55 60

130

Val Gly Gln Gly Gly Leu Glu Leu Leu Ala Ala Ser Gly Pro Pro Ser
65 70 75 80

Ser Ala Tyr Gln Ser Ala Val Ile Thr Gly Val Ser His His Ala Gln
85 90 95

Pro Leu Asn Ser Val Phe Tyr Ser Lys Ala Lys Ala His Val Phe
100 105 110

<210> 208
<211> 81
<212> PRT
<213> Homo sapien

<400> 208

Met Leu Ala Leu Phe Val Val Gly Gly Cys Pro Cys Ser Phe Gln Tyr
1 5 10 15

Met Arg Gly Gln Gly Asp Pro Arg Gly Pro Phe Cys Gly Pro Leu Trp
20 25 30

Lys Lys Gly Arg Arg Tyr Val Ser Cys Leu Ile Thr Ser Ile Lys Pro
35 40 45

Val Ala Cys Ile Ser Leu Lys Cys Ala Ile Tyr Ala Gly Ser Ser Gly
50 55 60

Gly Val Ile Tyr Val Trp Ala Pro Pro Arg Ala Pro Asn Thr Pro Leu
65 70 75 80

Tyr

<210> 209
<211> 67
<212> PRT
<213> Homo sapien

<400> 209

Met Lys Val Pro His Gln Arg Lys Lys Asn Lys Asn Thr Lys Lys Arg
1 5 10 15

Lys Lys Lys Lys Lys Val Leu Trp Gly Gly Tyr Thr Thr Cys Gly His
20 25 30

Asn Ile Gly Val Leu Pro Gly Val Cys Cys Ala Arg Thr Thr Trp Cys

131

35

40

45

Cys Val Ile Ile Thr Gly Gly Phe Ser Asp Lys Phe Phe Arg Asp Lys
 50 55 60

Lys Asn Leu
 65

<210> 210
 <211> 80
 <212> PRT
 <213> Homo sapien

<400> 210

Met Phe Met Cys Ile Cys Tyr Leu Pro Asn Tyr Ile Thr Ser Ser Leu
 1 5 10 15

Lys Val Glu Met Ser Met Glu Thr Asp Asn Met Ser Gly Leu Leu Leu
 20 25 30

His Thr Leu Gln Val Ser Ala His Leu Ile Phe Ile Ala Thr Leu Arg
 35 40 45

Asn Ser His Cys Tyr Pro His Phe Ile Ser Arg Gln Gly Lys Val Lys
 50 55 60

Ser Gly Lys Val Tyr Leu Trp His Lys Leu Leu Asn Glu Gly Thr Tyr
 65 70 75 80

<210> 211
 <211> 125
 <212> PRT
 <213> Homo sapien

<400> 211

Met Ser Ser Glu Val Ser Val Trp Glu Phe Val Gly Ala Gly Gly Leu
 1 5 10 15

His Gln Ser Val Ser Lys Gln Pro Arg Gly Lys Ala Lys Pro Leu Val
 20 25 30

Gly Asn Pro Tyr Trp Ser Phe Asn Arg Leu Ser Lys Gly Leu Phe Trp
 35 40 45

Lys Trp Glu Lys Ala Cys Cys Leu Pro Thr Gly Gly Glu Thr Thr Val
 50 55 60

132

Phe Gly Gly Leu Phe Pro Lys Leu Val Ser Lys Gly Asn Cys Trp Phe
65 70 75 80

Pro Val Phe Gln Lys Gly Asn Gly Phe Ser Val Ser Gly Trp Gly Ser
85 90 95

Asn Pro Val Leu Val Leu Gly Gly Val Asn Pro Arg Pro Lys Lys Ile
100 105 110

Lys Leu Glu Thr Ser Pro Tyr Thr Ala Lys Ser Trp Gly
115 120 125

<210> 212
<211> 167
<212> PRT
<213> Homo sapien

<400> 212

Met Arg Thr Trp Trp Cys Arg Val Leu Glu Val Arg His Val Ala Lys
1 5 10 15

Gly Gly Ala Pro Leu Arg Leu Arg Phe Leu Trp Arg Ser Val Ser Pro
20 25 30

Ala Cys Arg Glu Lys Glu Ile Ser Leu Ala Gln Thr His Asn Thr Arg
35 40 45

Met Arg Thr His Asn Leu Lys Asp Tyr Lys Arg Lys Ser Leu Arg Arg
50 55 60

Asn Asn Leu Leu Arg Ala Ala Ala His Ser His Val Leu Trp Arg Val
65 70 75 80

Ser Pro Thr Tyr Ser His His His Thr Met Cys Ala Val Thr Arg Cys
85 90 95

Thr Pro Arg Gly Val Leu Pro Ser Arg Gly Ser Ser Arg Val Cys Val
100 105 110

Lys Arg Ala Thr His Arg Phe Arg Cys Ile Leu Tyr Ser Glu Asp Leu
115 120 125

Trp Val Phe Ile His Ser Val Val Ser Ile Pro Phe Val Pro Val Gly
130 135 140

133

Val Lys Ile Trp Leu Pro Ala Leu Thr Ile Leu Pro Thr Thr Cys Gly
 145 150 155 160

Thr Lys Asp Thr Pro Leu Phe
 165

<210> 213
 <211> 151
 <212> PRT
 <213> Homo sapien

<400> 213

Met His Ala Arg Ala Ala Gln Cys Asp Gly Phe Ala Ala Arg Ser Pro
 1 5 10 15

Pro Phe Phe Phe Phe Phe Phe Phe Phe Phe Leu Gly Arg Gly Lys Asn
 20 25 30

Phe Phe Phe Phe Phe Ile Phe Ser Gln Lys Pro Phe Phe Trp Lys Lys
 35 40 45

Leu Lys Val Ala Met Arg Gly Phe Leu Tyr Lys Lys Asn Ile Lys Thr
 50 55 60

Arg Gly Ile Leu Leu Phe Gln Lys Lys Phe Asn Leu Leu Phe Val Asp
 65 70 75 80

Lys Ala His His Glu Trp Val Tyr Lys Leu Val Leu Ser Tyr Ile Phe
 85 90 95

Gln Arg Lys Tyr Tyr Ser His Ser Val His Val Tyr Ser Ile Thr Val
 100 105 110

Cys Ser Arg Arg Lys Ser Arg Arg Ala Cys Asn Ser Leu Gly Val His
 115 120 125

Lys Cys Val Leu Pro Leu Cys Glu Ile Leu Cys Phe Ile Pro Val Pro
 130 135 140

Gln Tyr Ser His Asn Asn Ile
 145 150

<210> 214
 <211> 118

134

<212> PRT

<213> Homo sapien

<400> 214

Met Leu Cys Arg Ser Val Cys Asp Tyr Pro Pro Ala Arg Val Arg Arg
 1 5 10 15

Glu Val Val Val Cys Asn Thr Lys Arg Gly Gly Gly Arg Arg Arg Glu
 20 25 30

Gln Pro Ser Ile Thr Arg Val Ala Ala Leu Ile Tyr Ile Tyr Met Val
 35 40 45

Glu Gly Glu Ile Lys His Ile Ser Arg Glu Arg Glu Gly Glu Arg Ala
 50 55 60

Asn Pro Thr Thr Ala Gly Gln Gln Glu Ala Ile Ser Arg Gly Glu Glu
 65 70 75 80

Glu Arg Gly Cys Ser Ala Arg Arg Ala Pro Thr Pro Pro His Asn Thr
 85 90 95

Leu Tyr Arg Thr Gln Gln Thr Lys Pro Gln Pro Arg Thr Gln Ser Thr
 100 105 110

Arg Glu Tyr Lys Lys Ile
 115

<210> 215

<211> 72

<212> PRT

<213> Homo sapien

<400> 215

Met Val Ala Met Ile Ile Arg Ser Ile Phe Val Gly Leu Leu Ala His
 1 5 10 15

Ser Cys Cys His Ala Gly Asp Asp Thr Phe Arg Ala Pro Leu Ala Leu
 20 25 30

Ile Leu Glu Leu Leu His Leu Ile Val Val Gly Phe Trp Asp Ser Val
 35 40 45

Ser Val His Ile Asp Thr Pro Pro Glu Glu Leu Leu Met Ile Phe Phe
 50 55 60

135

Leu Gln Gln Cys Ser Tyr Val Val
65 70

<210> 216
<211> 58
<212> PRT
<213> Homo sapien

<400> 216

Met Cys His Cys Pro Arg Val Pro Pro Ile Pro Gln Ala Thr Asn Phe
1 5 10 15

Val Thr Arg Glu Gln Ile Gln Glu Ile Ser Ser Gln Ala Lys Val Gln
20 25 30

Ser Ala Ala Asn His Gly Arg His Ala Glu Pro Arg Arg Arg Cys Ala
35 40 45

Ser Leu Val Pro Gly Ser Asp Gly Ala Ala
50 55

<210> 217
<211> 121
<212> PRT
<213> Homo sapien

<400> 217

Met Gly Gln Asn Gly Val Ser Pro Gly Gly Lys Cys Gly Cys Thr Gly
1 5 10 15

Leu Lys Ile Pro Thr Lys Gln Phe Glu Thr Thr Lys Asn Glu Gln Gln
20 25 30

Gln Glu Lys Glu Glu Gln Thr Arg His Thr Arg Asn Arg Arg Arg Arg
35 40 45

Glu Arg Glu Arg Asn Thr Asn Thr Gln Gln Pro Arg Lys Asp Glu Lys
50 55 60

Glu Arg Glu Lys Arg Glu Arg Lys Glu Glu Lys Arg Glu Asn Lys Lys
65 70 75 80

Lys Glu His Gln Lys Glu Lys Lys Asn Thr Lys Thr Arg Gln His Thr
85 90 95

136

Lys Gln Arg Lys Thr Gly Arg Thr Thr Lys Glu Asp Lys Asn Ser Asn
100 105 110

Glu Lys Gln Glu Arg Thr Lys Thr Lys
115 120

<210> 218
<211> 67
<212> PRT
<213> Homo sapien

<400> 218

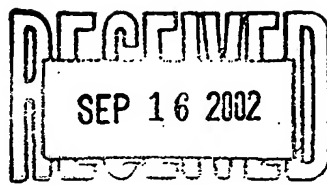
Gly Pro Gln Gly Pro Pro Gly Tyr Gly Lys Met Gly Ala Thr Gly Pro
1 5 10 15

Met Gly Gln Gln Gly Ile Pro Gly Ile Pro Gly Pro Pro Gly Pro Met
20 25 30

Gly Gln Pro Gly Lys Ala Gly His Cys Asn Pro Ser Asp Cys Phe Gly
35 40 45

Ala Met Pro Met Glu Gln Gln Tyr Pro Pro Met Lys Thr Met Lys Gly
50 55 60

Pro Phe Gly
65



PATENT COOPERATION TREATY

PCT

From the INTERNATIONAL BUREAU

To:

LICATA, Jane, Massey
Licata & Tyrrell P.C.
66 E. Main Street
Marlton, NJ 08053
ETATS-UNIS D'AMERIQUE

Docket System _____
Status Report _____
Docket Book _____

NOTICE INFORMING THE APPLICANT OF THE
COMMUNICATION OF THE INTERNATIONAL
APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

Date of mailing (day/month/year) 06 September 2002 (06.09.02)		IMPORTANT NOTICE	
Applicant's or agent's file reference DEX-0268			
International application No. PCT/US01/45151	International filing date (day/month/year) 20 November 2001 (20.11.01)	Priority date (day/month/year) 20 November 2000 (20.11.00)	
Applicant DIADEXUS, INC. et al			

1. Notice is hereby given that the International Bureau has **communicated**, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this notice:
KP,KR,US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:

AE,AG,AL,AM,AP,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EA,EC,
EE,EP,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KZ,LC,LK,LR,LS,LT,LU,LV,MA,
MD,MG,MK,MN,MW,MX,MZ,NO,NZ,OA,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,
UZ,VN,YU,ZA,ZW

The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this notice is a copy of the international application as published by the International Bureau on

06 September 2002 (06.09.02) under No. WO 02/068645

4. **TIME LIMITS** for filing a demand for international preliminary examination and for entry into national phase

The applicable time limit for entering the national phase will, **subject to what is said in the following paragraph** be **30 MONTHS** from the priority date, not only in respect of any elected Office if a demand for international preliminary examination is filed before the expiration of **19 months** from the priority date, but also in respect of any designated Office, in the absence of filing of such demand, where Article 22(1) as modified with effect from 1 April 2002 applies in respect of the designated Office. For further details, see PCT Gazette No.44/2001 of 1 November 2001, pages 19926, 19932 and 19934, as well as the PCT Newsletter, October and November 2001 and February 2002 issues.

In practice, **time limits other than the 30-month time limit** will continue to apply, for various periods of time, in respect of certain designated or elected Offices. For **regular updates on the applicable time limits** (20,21,30 or 31 months, or other time limit), Office by Office, refer to the PCT Gazette, the PCT Newsletter and the PCT Applicant's Guide, Volume II, National Chapters, all available from WIPO's Internet site, at <http://www.wipo.int/pct/en/index.html>.

For filing a **demand for international preliminary examination**, see the PCT Applicant's Guide, Volume I/A, Chapter IX. Only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination (at present, all PCT Contracting States are bound by Chapter II.)

It is the applicant's **sole responsibility** to monitor all these limits.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland Facsimile No. (41-22) 740.14.35	Authorized officer J. Zahra Telephone No. (41-22) 338.91.11
--	---

